A Partial History of Emotions and Some Potentials and Problems Surrounding the Exploration of Artificial Emotions

Interview between Arvid Kappas and Bill Seaman – Spring 2013

Key Words – emotions, emotional categories, dimensional approaches to emotion, artificial emotions, machinic emotions,

Seaman – Arvid, to begin I am wondering if you could give a brief history of the field of Emotion research and later we will discuss how this ties into the production of artificial emotions in robotics. Are there a specific set of emotions that are universally agreed upon or should this be a more nuanced question?

Kappas - There are some questions regarding what basic emotions are, and things related to that issue. There are things that can be easily distinguished one from another like apples and bananas and they are biologically easy to distinguish because there are many ways in which they are different, so these are “natural” kinds… this means in the world there are objective properties of these things that make them different. Now the question is ‘are emotions natural kinds or not’? because if one asks if there are basic emotions for some authors there is this idea that there is a small group of emotions like happiness, sadness, anger, fear and surprise which are different from each other; which have biological origins; and because of that in a way are as different from each other as bananas and pears. But not everybody shares that point of view. It is a very important point of view and the person who is most identified with that notion today is Paul Ekman. So he is probably the most important living emotion researcher today. He got famous in the late 60s by doing a couple of groundbreaking studies and by being involved in the development of a tool which is called FACS – Facial Action Coding System, that he published together with his colleague Wallace Friesen in 1978 and which is a way of describing objectively facial behavior based on an anatomical description of what the face can and can not do. That has had a huge impact on the field because it allowed to test some of the assumptions using verifiable, repeatable research. Ekman is inspired by the work of Charles Darwin. Basically, I will often make the statement that while people have thought about emotions for thousands of years, if I had to pinpoint a critical event it would be the publication of Darwin’s book The Expression of the Emotions in Man and Animals in 1872. I see this as a year zero for the scientific research on emotions. You can always argue about that because that involves evaluating the contribution of this person or that person, but for me this is really the key thing. Ekman was one of the people who was very much inspired by this book and who has emphasized that fact many times. So while there were a couple of re-editions of Darwin’s book, the latest in the 1950s, Ekman re-edited this in the 90s and provided a version of the Darwin book with comments and photos from Ekman’s own research, to try to do … ‘this is what we know today,’ together. Some people have found this to be inappropriate, because unless you are a scholar of the field it will not always be obvious what is Darwin and what is Ekman, because for example on the cover he blends illustrations that Darwin would have used with illustrations that he used. People said this is him putting himself into this thing. So, having said that, one of these important studies that I mentioned was published in 1969 in the Journal “Science”, which of course creates immediately much credibility and according to Darwin you would expect that there are expressions of certain emotions that should be universal, because emotions and the expressions of the emotions are embedded in evolutionary history. People were very excited when Darwin made these arguments, however, when people started looking for this they failed to find easily this universality. They would say "I have difficulties understanding people in my own culture, so how can it be that all over the world it is like that". So Ekman actually did a study where he took photographs of people expressing what he thought to be basic emotions, showed them in different countries and most spectacularly a tribe in New Guinea, the Fore, where these people had no experience with westerners whatsoever, they had to have a translator with a similar language to communicate with them and they would recognize the American expressions. And also he said to them, I will tell you a scenario… imagine you come out of your hut and there is a dangerous animal – a wild boar that might attack you, what would you show on your face, took a photo of that, and brought these photos back and the Americans for example were able to recognize these emotions. So this study was taken as proof that there are certain emotions that are recognized all over the world.

Seaman – Do you believe him? Was this an empirical study?

Kappas – It has been criticized a lot and one of the biggest critics in this sense is somebody called James Russell who published a paper in the late 90s criticizing these studies for many reasons. Between Ekman and Russell there was a fierce exchange of publications

Seaman – Of emotions (laughing)

Kappas – Of emotions, to the point were according to rumor a law suit was threatened actually. It turns out that not all things happened the way they were being told. There were not studies in the strict sense, where people were talking together back and forth and then they came up with explanations of what this could be or what that could be.

So, yes the study has been criticized also.

[[can we use this now – it wasn’t published at the time and you wanted to leave it out]]

Ekman was influenced by Sylvan Tomkins who was a genius, but who was also weird in some ways. He wrote these monster books on emotions in the early 60s.

According to Tomkins emotions are in the face, but he argued that there are more than the 6 emotions –

Seaman – How many did he suggest?

Kappas – I don’t know by heart --- something like 10… and from Tomkins comes this notion that there are affect programs in our head – these are like embedded things in the brain that tell what the emotion feels like, what it looks like, what the bodily responses are. Basically it is like a module in your brain and when it gets “kicked”… when the whole thing starts.

Seaman – Is this similar to innate releasing mechanisms of Konrad Lorenz?

Kappas – It is related to that. And the whole idea that Ekman had was a derivative of Tomkins' ‘Affect Programs’ because Tomkins like Darwin thought of more than those six, Ekman happened to have a sufficient number of photos of men, women and children, expressing these emotions. Possibly there could have been more than six included at the time.

Ekman was inspired by Tomkins who actually had more than those six emotions and after the fact they added a seventh one which was ‘contempt’, which might be a universal emotion…

Seaman- Can you list them again slowly here?

Kappas – Happiness (the only positive one), Anger, Fear, Sadness, Surprise, Disgust

Seaman – Contempt is seven.

Kappas – However in some studies disgust and contempt are easily confused. There is another researcher that was at about the same time influenced by Tomkins. This guy is called Carroll Izard, and he has 10. He includes Interest for example which has also to do with the fact that Izard tends to work mostly with children. He has this belief that you can get more about emotions by looking at children because they are less regulated and they are more spontaneous.

So if you would say you take the triad of Tomkins, Izard and Ekman, you have the prototypical theorists related to the basic emotions. They are clearly not the only ones but if you would do a distribution of researchers in a cloud, they would be pretty much at the center.

Seaman – Do you see these as separate or as force vectors that get mixed?

Kappas – Well some people would say this is absolutely silly, if you do a list of emotions words in the English language you come up with thousands of words so how do these come about --- are they different things. Some would argue, I believe Izard would and to a certain degree Ekman, that these are blends. So for example jealousy would be not a basic emotion but it would be a mix of fear of losing your partner combined with anger about a behavior that you observe or that you come to know.

This is now stuff that dates back to 1960, when the Tomkins books were published, I think 1962 and 63. The late 1960s which is when Ekman’s stuff was published, and then much work was done in the 1970s. Then I would say a new important branch was opened in 1980 when a guy called Robert Plutchik published a book called The Psycho-Evolutionary Theory of Emotions. That relates to what some of these people did. It also has a strong evolutionary perspective, but he deals also with states that are also not those basic emotions. He has a 3 dimensional model. I think there are 8 states. There is a sort of cone that relates to these and he links this to momentary states as well as personality differences. What is nice about this Plutchik book is that the first part of it is like a review of emotion research up to that point, going back to William James cannon but also some of the people who are often not mentioned in this context like the behaviorists or Freud and other psycho-dynamic ideas. That is a useful thing.

As a side remark, some of the things I am talking about now can also be found in the textbook that I use for my Emotion Course which I like a lot because it creates a nice historical context and this is titled The Science of Emotion by Randy Cornelius, 1996. It is a relatively small book. What I like about what Cornelius does is the way that he chops up the complicated story into saying if we take all of these different theories and there are many together we can lump them in four big schools of thought: one is the Darwinian tradition (this relates to what I was just talking about basically), the Darwinian tradition has much to do with the face and the expression of basic emotions; the second on is the Jamesian tradition and that goes back to William James who comes from a different angle — for him physiology is much more important and how physiology affects how we feel.

Seaman – It's the whole body…

Kappas – Yes for him it is the whole body and it is not that there is a program as such, that what we feel is not that the program triggers it, but something triggers an emotional response like you meet the bear in the woods, you start to run… and all of that, the behavior of running, how that feels, your heart with beating, this is how we feel fear, and if there was no ‘bodily’ response we would not feel fear.

Seaman – If it is not a program then, it is just purely perturbation?

Kappas – The idea for him was that there were many many different ways in how our body might respond and this has this strong impact on what an emotion feel like. Now James was an armchair researcher, he would sit down and think very hard about this.

Seaman – Its more theoretical.

Kappas – Yes, and because of that it is also something that had a long impact, because we can still today read these papers and relate to that very well.

Seaman – Are you closer to that theory yourself?

Kappas – Ummmm, (smiling) I don’t know yet… I’ll tell you at the end.

Seaman – (Laughter)

Kappas – Curiously enough, James has become very popular these days. When I have gone to conferences in the last few years, one of the big topics that comes up is embodiment. It is a response to the cognitive wave which we haven’t touched on so far, because that comes after the 60s and 70s. For people, cognition became this mantra that explains everything, and today people argue that it is cognition, disembodied, doesn’t get you anywhere. You need the body. You need what the body feels like. If you put somebody in a situation different things happen then when they just think about it. This is of course close to what James had originally said. Its like what your body ‘does’. There are all kinds of flashy experiments that people talk about these days, a for example very simple thing: imagine that we go to the outside where Jacobs University is a green, and the corners of the green are not flat, but they have a slope. And I ask you to estimate that slope. It seems easy enough. You might be good, you might be bad, at that. You say a number or you indicate it with an angle. Parallel universe. We do the same thing, but I put a backpack on you that is full of heavy stuff. Experiments show that with a heavy backpack it seems steeper. This is embodiment. This means that a physical estimation is not just a perceptual problem, but somehow your body resources play a role.

Seaman – Historical.

Kappas – I give you something to drink. In one scenario I give you an energy rich sugary drink, in the other condition I give you a diet drink. You don’t know whether it is the diet drink or the sugary drink. People who have the sugary drink estimate the slope as less steep.

Seaman – Interesting.

Kappas – Because you body feel energetic.

Kappas – Different experiment. We bring a friend along. If you are together with a friend the slop looks less steep. So these are a whole series of things.

Seaman – So this is environmental.

Kappas – yes.

Seaman – Environmental, and Historical, and is it even social and cultural?

Kappas – Yes. And so in the last few years one book that has influenced many people outside of academia is Damasio’s book Descartes' Error. There is a second book after that, The Feeling of What Happens by Damasio --- the two are related. Damasio is a Neo-Jamesian in the age of Neuroscience. So much of the theory is Jamesian with the twist that now we actually have some knowledge about what actually happens in the brain. But it is very very close. So this is what I mean… despite the fact that James wrote these things at the end of the 19th century, they are again very popular these days.

Kappas – OK. We were talking about the schools of thought according to Cornelius.

The first one was Darwinian; the second one — the Jamesian; the third one is the Cognitive tradition and with that we enter mainly the 80s. While the starting point was also in the 60s with the work of Magda Arnold and Richard Lazarus it was a wave of new researchers, new scientists like Nico Frijda, Klaus Scherer, Craig Smith, and Ira Roseman.

Seaman - And you were working with Scherer correct?

Kappas – yes

Kappas -- …and Ira Roseman who explored further how cognitive processes would trigger specific emotions. So according to James, you see the bear and you run…Why? Why do you run when you see the bear? If you read James there is sort of a direct connection in the brain, while he didn’t know better — that doesn’t work well. So somewhere your brain must have made a decision that there is a potential threat, and that you need to get out of there.

Seaman – Doesn’t this go back to the oldest portion of the brain…

Kappas – So these cognitive theories became known as “appraisal” theories, basically answered this question, how do different states come about. So somebody like Scherer who thinks that we ‘appraise’ our environment along several dimensions… you end up with a multi-dimensional space. That has truly has endless… limitless numbers of states. So rather than having 6, 7, 8 or 10 basic emotions, for him there is an infinite number of emotions.

Seaman – This go back to Kurt Lewin’s Topological Psychology?

Kappas – I am sure that it is influenced by that.

Kappas – So how does he reconcile with the fact that if you ask people they seem to come up with a basic number --- a small number of emotions, Scherer says – well, where there is the possibility of infinite combinations, not all combinations are equally probable. Some of these combinations are bound to occur more frequently. In a multi-dimensional space, he argued that at some point we could think of the idea that there are some “strange attractors” and things converge to that. So that if you have something that is very close to… you still say this is anger. Even so, it is slightly different. And that is very important because, if there would be only 6, 7, 10 basic emotions, and if there were affect programs that would always lead to the same physiological responses, it should be easy to find these. But we don’t.

Seaman – Much more subtle.

Kappas – It is much more subtle and for Scherer it is easy to explain because he says, ‘anger’ does not equal “anger”. You might have 10 -100 different variants and of course physiologically it makes a big difference if somebody insults you at the schoolyard, or you realize this is the third time that somebody has done that. To him, one will relate to a very sudden response – very physical – you want to punch someone. The other would be this cold anger type thing where you plan revenge. His theory is particularly powerful in explaining these differences.

Seaman – Does he have a whole matrix that he draws?

Kappas – Yes – this all exists and a good way to get started into “appraisal” theory is a book which was published I think in 2001. It is called Appraisal Processes in Emotion: Theory, Methods, Research (Series in Affective Science), Scherer is the first editor, then Schorr, and Johnston is the third one, and that basically has a chapter by each of the key people including what I think is the last chapter by Lazarus before he died. There is Roseman, Frljda, there is Smith, of course Scherer and there is a bunch of critical chapters as well — where I have a chapter in there talking about in this case, the danger in using metaphors to far. The title is called – A Metaphor, Is A Metaphor, Is A Metaphor — Exorcising the Homunculus from Appraisal Theory (or something like that). That book has all you need to know including a decent summary of Scherer’s theory and some of the empirical stuff.

Seaman – Excellent. Just as an aside, this approach is very close to how Rössler and I have discussed how we would work on surrogate emotions— many force vectors and they would sum in different ways.

Kappas – It is a very attractive way and when I studied with Scherer, my Master’s degree was in 1986 and what I wanted to do for my PhD was an AI (Artificial Intelligence) model of Scherer’s theory, because at the time AI was big— expert systems, production systems, with languages like OPS5 or LISP based systems.

Seaman – This was 86?… that was very early on!

Kappas – That was my plan and I had a little proposal. I had learned some LISP and PROLOG during my studies

Seaman – Can I get a copy of that?

Kappas – I have no copy. It’s long gone. I even wrote a BASIC program that simulated what it should look like. So according to Scherer, we constantly evaluate our environment. The way he describes it is like a radar… there is the gesture of a rotating thing that constantly scans… a fixed sequence in his theory — that you check whether something new happens, if it does whether it is pleasant or unpleasant, whether it relates to your goals, whether it helps or hinders your goals, whether you can cope with it, and how other people would relate to that. And this sequence again and again would happen. The idea would be that if something happens that is new, your brain will reconfigure— you will try to put attention because it might be potentially relevant. So there are things you can measure with Event Related Potentials of the brain, maybe skin conductance, however, what occurred to me at the time was that we did not have the knowledge to predict how all of that activity would add up and overlap. Because it gets very complicated.

Seaman – It is an emergent model --- would you call it that?

Kappas – Yes. And imagine that something happens, there is a loud sound, you turn around, you see a dark shape, a moment later you recognize who it is, but then you see something else moving rapidly, so if each of these events is evaluated and causes a cascade of changes in your body and your brain, these will overlap, and if we don’t know the time causes of these events, I could not predict, if I take a slice out of that at any given moment, what the body should look like if the theory is right, because you have all of that lingering stuff of the previous events. So my thought was — let’s do an AI model where these individual processes are simulated with their own time courses.

Seaman – Do you see yourself as being one of the first people proposing this temporal model?

Kappas – No because the temporal model is implicit in Scherer's theory, for sure. All I did at the time… I worried how could we actually use this and test it and how a computer model would help in doing that. So at that time a friend of mine, Kim Silverman, who had worked with Scherer, he left Giessen where this all happened, and went to Cambridge to do his PhD, there in England. He did his work on synthesizing speech with Anne Cutler at the Applied Psychology Unit of the Medical Research Council and at that research institute there was the creme de la creme of the British cognitive researchers at the time. Many of which who were interested in AI and so forth. Kim said – well why don’t you come over and talk to some people here about your ideas. Which I did and they all said this was crazy. Mainly because we don’t know enough. We don’t know enough about the physiology for example to model these things. So it became obvious that we need to know more… that “I” need to know more. So I decided that before pursuing doing this PhD, I would have to get more knowledge in two fields mainly Artificial Intelligence and the psychophysiology of emotions. And this is how I got to Dartmouth. Scherer said he had been to Dartmouth College a bit earlier and there were people working on the psychophysiology of emotions and they had a strong AI.

Seaman – It was born there.

Kappas – The term AI was coined by John Mccarthy at a conference at Dartmouth in 55; LISP was born there. So that seemed like a good idea. I got the money to go there for a year. I ended up with somebody named John Lanzetta who was a social psychologist. John was unusual as a social psychologist because he had training as an engineer. So he was very much influenced by things like Lewin, and so forth. And he was a social psychologist which means for him, all of these processes that we talk about make sense between people. But they do not make sense if you just look at the individual.

Seaman – you can’t isolate it, it is always a cultural milieu.

Kappas – Everything we have talked about so far is very person-centered. But Lanzetta was very much interested in how interaction between people affects these things. So the first thing that John did was — was say – "well there is a course I think you should take. Systems’ dynamics". So I took a course on systems’ dynamics there, and the background to that is in the 1970s there was this book published on the limits of growth – Jay W. Forrester at MIT . and one of the people from the group was called Dennis Meadows. Meadows went to Dartmouth and founded there this group who was looking at ways of modeling dynamic systems. When I went there, in 1986… and the Macintosh was 84. So when I came there everything was already Macintosh at Dartmouth. And the team of Dennis Meadows had created a tool on the Macintosh that at the time was called “Stella” and that allowed you to create conceptually a dynamic system… then you put some values into it and you let it run over time. So I actually ended up as the term project to do something with Scherer’s model in that environment.

Seaman – Just a slight aside. Does this model from Lanzetta tie into the Linguistic as well.

Kappas – Not so much. It is more about Lanzetta who was influential in many ways. For practical reasons he was editor for several years of the Journal for Personality and Social Psychology which is a very important journal in the field. And he did some early research on the so-called facial feedback hypothesis.

Seaman – Related to Mirror Neurons…

Kappas – Before there were Mirror neurons so to speak. Mimicry, empathy— these were very important things. He did this study together with some colleagues there that is still very very important, which is still very much overlooked by this mirror neuron discussion. At the time Lanzetta, his colleague Bob Kleck also a psychologist and two political scientists, Roger Masters and Dennis … [[tip of the tongue… did some studies on how people respond to politicians So this was the time of Ronald Reagan. The democratic challenger was Hart. Who later was sidelined by a scandal. So what they did was, they had people looking at these videos and they were taping all of those for years basically and if it was all mirror neurons you would expect that people would smile when the see somebody smile. But what they could show was that Reagan supporters smiled when Reagan smiled and frowned when Reagan frowned. But the Democrats would frown when Reagan smiled.

Seaman — Social and environmental influences…

Kappas – Yes. So this is one of the important lessons from this line of research. That the social context also determines things like empathic and counter empathic responses.

Short break

Kappas – We have been talking with a sideline about some of my own history in this story and we arrived in Dartmouth talking about Lanzetta and Kleck.

Seaman – So one year at Dartmouth for you?

Kappas – As it turns out I also took a grad course in AI for building Expert Systems and I took a grad course in physiology and modeling physiological systems so all that I needed then I decided to stay for two more years. And abandoned the idea of going back to do the other model because it became obvious that it was too early to model Scherer’s theory. Too little was known with regards to how these physiological systems behave.

Seaman – Although it is kind of interesting in trying to make the model because you have to start sometime…

Kappas – Of Course, However, I was very much then drawn into the social aspect and looking at emotions from a social element. Indeed, something that these other theories and theorists often tended to overlook— the social component. From my point of view, this is where I see much of my own thrust today. And that has to do with emphasizing how important the social aspect of emotions is. So that we should not look when we try to understand emotions, or model it the way that you want to do, to model a dismembered brain that is floating in empty space, but to realize that not only are we floating in the space with connections to each other, but that our brain evolved to do that in a very effective way. Our brain is social. There is this new field called Social Neuroscience that makes much of this, and the 'prophet' of that thing is John Cacioppo from the University of Chicago. He published an important paper in 1992 in The American Psychologist together with his colleague Gary Berntson on “The Decade of the Brain”. The term ‘Social Neuroscience’ was basically coined in that paper. It was a sideline thing. The point of the paper was that brain research is important in the light of the decade of the brain, that the Bush administration and Congress had initiated. That in order to make sense of the brain, we need to connect the brain to different levels of analysis. The paper by Cacioppo and Berntson was on multi-level analysis. This was truly a programmatic manifesto type of paper that now, two decades later, is really coming to fruition.

Seaman – So before this paper this notion of a multi-level analysis wasn’t the main emphasis.

Kappas – It was implicit. Cacioppo himself is a psychophysiologist and within psychophysiology people have been talking about these things for longer, but this paper brought it to the front (in a way). For me the social context means these are some levels, in this multi-level enterprise that are neglected.

Seaman – One of the things that came out in my conversations is that there are many people working on these separate levels, but there is very little conversation by these researchers between the levels or bridging the levels. For me it is this long-term goal, of looking at the relation between these levels.

Kappas – What I would suggest is that you go to Cacioppo’s web page at the University of Chicago, almost all of his publications are there. There are several which have titles like “Bridging Levels”. You will find easy material which does not require you to be a specialist in physiology or neuroscience.

Kappas – Just to close the chapter then, in the Cornelius book, the fourth school is social constructivism. However this is not exactly the same type of social stuff that we have been talking about for the last few minutes, but basically he relates to people who discuss to what degree, what emotions are, are actually cultural constructs. This fits into a larger theme of Social Constructivism where topics like Gender, what it means to be human, culture, all of these constructs… however it is really difficult at this point for Science to pretend that there are no biological constraints. So the reality is somewhere between the extreme Social Constructivist position and the biologically oriented ‘basic emotions’ tradition that we started with Ekman, Izard and Tomkins — it is somewhere in-between.

Seaman – Lets see if I can sum it up – there is bio physiology, the social, the cultural, the temporal, and the environmental.

Kappas – That is a good starting point.

Seaman – So each of those are also vectors in a sense.

Kappas – And they are not independent vectors.

Seaman – They are intra-active and accretive. And it is also a learning environment.

Kappas – That is for example part of the environmental.

Seaman – and it is also embodied.

Kappas – Yes.

Kappas - So in the early 80s a handful of emotion theorists, frustrated with the state of the field said emotion is something which is very difficult to get a hold of. You need different backgrounds — not only psychologists, you do need philosophers… different countries have different research traditions, it should not be American psychologists who come up with a story that works only there… and they founded a society, called The International Society for Research on Emotions – ISRE

Kappas – So if Neosentience gets into the emotion field, you want to get to ISRE

You want to go to this conference.

Seaman – And give a paper.

Kappas – Because it is a highly interdisciplinary gathering. The next conference will be next year in late summer in Berkeley, California. The last one was last year in Kyoto, Japan.

Seaman – Is there a proceedings?

Kappas – Yes, for some of them.

Kappas – As it turns out among these handful of people who founded the organization was Scherer and Ekman. I was there at the time when all of that happened, and got sucked into this.

Seaman – Are you going to bring it here.

Kappas – I did organize it in 2000 in Quebec City. It is a once in a lifetime thing.

Seaman – It is a huge amount of work.

Kappas – The Society now has a Journal which is called “Emotion Review” which is published at Sage.

End 1st meeting.

Kappas – In the first session we talked briefly about historical approaches to emotion, we had the Darwinians, Darwin himself and then people influenced by him — Tomkins in the 1960s who influenced Izard and Ekman and this is where the idea of basic emotions, I wouldn’t say comes from, but this is what people would pick out of a hat if it came to the issue of that. And we talked about some other views— physiology related, cognition related, and socially related, so now we will jump to a parallel universe.

Kappas – As you know when physicists look at light they found out that there is some behavior of light that is consistent with the idea that light consists of waves, and other phenomena… that it consist of particles and this is not the same thing. So this is quite confusing in a way because you have to assume that it has to be two things at the same time. The same is true when looking at emotions. So far we talked about categories of emotion, this creates emotion states. However, there is a long tradition that is currently getting more attention that looks at dimensions underlying emotions. So rather than thinking of emotions as discrete entities, they would say there is an n-dimensional space and while it is clear that this is like a different level of abstraction, there are certain phenomena that just make more sense when we think of it in terms of dimensions. For example, let’s say somebody annoys you and you are angry and then the person is more annoying so OK, is it the same anger or is it a different anger because it is at a higher intensity. What if you suddenly realized that there is something about what the person does that was not intended. But you are still angry. Is it the same anger? And so forth. In a dimensional space you don’t need to obsess about how you are going to call it, you just think of it as a point in the space that is moving around. So it is particularly useful for applications where you try to generate something. But also where you try to analyze the emotional state of something. You can find everything with Greek Philosophers if you go looking for it. I will not go that far but I will start, just as I did with the discrete emotions with Darwin, with Wilhem Wundt, the German founder of Psychology who developed a complete theory about three fundamental dimensions of emotions. And this in one way or the other, if it is attributed to Wundt or not, has been appearing at different points of time in the history of science — some people along the way would be Harold Schlosberg who looked at how facial expressions could be seen as representing two or three dimensions. And very important for many fields— Charles Osgood who invented the semantic differential in the 1950s and basically the idea that you can measure and compare anything along these dimensions so you can for example take a bottle of coke, take the American President, or a record that you like. You measure them along these dimensions and you see how close they are in space. This is something that has influenced people from advertising and what not. In the more recent history, the primary dimension that most people agree on is simply valence, positive / negative. Everything that we do or perceive somehow relates to positive / negative— whether we like it / whether we don’t like it; whether we want to approach it or whether we want to get away from it. This is so fundamental that there is basically no organism on this planet that does not react to that. So you can take a simple, few cell complex organism and it will approach things where you can eat something or if it gets too hot it will go away. It can actually learn also to avoid things or to search for things, according to things being good or bad. There are some people who believe that when a baby is born, basically the only thing that is hard wired is its sense of pleasure and the sense of pain and everything starts to unfold from there. You do not need to learn pain. This is something hardwired which is present at birth. Similarly if you take sugar water, you can take an infant which is a few hours old, drop that on the tongue and the infant will smile. So this is really something that is hardwired. Now about 100 years ago in the US psychology started to be dominated by Behaviorism, more so in the 20s. Then going down in the 40s. The key aspects of Behaviorism are the learning mechanisms— the conditioning mechanisms that at the same time in Russia, Pavlov was interested in. Pavlov worked with classical conditioning and the Americans were more interested in things like operant conditioning, however it always has to do with something that reinforces a particular thing— something that is positive or negative. So if you take Pavlov’s Dog, you have the food that the dog likes. The dog reacts to the food by salivating for example. So you pair a neutral stimulus, like a bell, with the food and then the bell will elicit the salivation because the bell will start to be positive for the animal. This has been in more recent years looked at with something that is called affective conditioning. This is a variation. Where basically just by placing something that is neutral next to something that is positive, the neutral thing will start to trigger these reactions.

Kappas - In the brain we find much evidence that positive / negative, and approach/avoidance is very critical. We can look at facial expressions and we can see that there are certain movements like the eyebrows pulling together, that relates very well to how positive or negative something is. That has now been shown reliably by many researchers. So it is also being increasingly used as an indication of whether you like something or don’t like it, without people having to explain or fill out questionnaires. And without them perhaps even being aware of it. So one hot topic is for example to look at things like racism with so called implicit measures. So you might have somebody who says 'no, I am not racist," then you ask them some questions, there is no evidence. The next experimenter comes in and is black and you see that the eyebrows go together, the same way they would if you look at something you do not like. So people are interested in these things and in turn, the success with these measures reinforces the notion that it is useful to think of emotions in this dimensional space, because we can do so much more. There is a whole new field in computer linguistics which is called sentiment mining. Sentiment mining relates to getting the sense of the feeling that is implied, in text. What that looks at, typically, is not whether there is anger or happiness, or fear, but it looks for positive / negative. For example, Obama gives a speech and you go to Twitter, grab 15 minutes of twitter, you use the software that tries to identify positive/negative, and you have a pretty good sense of how people viewed that speech.

Kappas - The second dimension is a little bit more complicated. Typically this is arousal. The thing is that valence could be seen as something that relates to the object that you look at, so you look at a painting that you like and you find it pleasant, then people would say it is also a property of the object, the painting or whatever, but it is also how you feel about it. Now arousal seems much more linked to how you feel about it. It is your body being activated by something and typically that would be because it relates to something that requires action. So if you are late and you want to catch a bus, and you walk out of the gate of Jacobs University and you see the bus coming up there you get very activated because it is relevant and your body needs to get going. Arousal relates to mainly bodily states. People like James who said that you feel emotions basically because of bodily reactions. That is close to this arousal concept. However if you would say are these two dimensions equally important — no they are not. If you take people and you have them evaluate the world around them, then valence seems to be much more important than arousal. Also, despite the fact that many people believe that they are independent, there is reason to believe that they are not. In the sense that if something is extremely extremely negative, it is difficult to imagine that it would not also be arousing. So the people who study this today often used standardized images to see how people respond. There is a Center for the Study of Emotion and Attention in Florida, in Gainsville, which is headed by Peter Lang and Margaret Bradley. They have been working for quite some time there and one of the things that they did was the creation of a set of stimuli of images called the international Affective Picture System (IAPS) which now has over a thousand pictures which have been used in many many studies so you know if he looks at the bunny this is positive 3 in arousing children, and if somebody looks at the chopped off hand, then it is negative and so forth. As a matter of fact. We are using that in our research here and we have found, particularly for valence, almost perfect correlations between our subjects and the American subjects, there. So we have the valence, we have the arousal, and the third one is sometimes called dominance or sometimes called power— strong/weak, something like that. Imagine that somebody bumps against you and you are upset because the person doesn’t pay attention, it is negative and it is arousing. And because the guy looks at it in the same format as you, you get quite angry. It is negative, it is arousing, and you have a certain level of power. However, if you turn around and the guy is the big hulk, then low power— it is not the same feeling, this is fear. Fear would be a state that is characterized by negative, high arousing, and low power. So this is where this power dimension comes in. It is particularly important for the distinction of fear and anger.

I will now draw a coordinate system. I will call one valence, and the other one arousal, and as it turns out, if we map these images you do not get a round cloud, but you do get a sort of triangular shape that is sometimes referred to as a boomerang. And that has simply to do that it is very difficult to find high arousing images that have no valence. So if something is not positive, or not negative it is not very arousing. In most of these maps you do not see that third dimension. So if I would now draw 3 dimensions, I could imagine a ball, as a 3 dimensional thing but probably that is not what it looks like. It could be much more like a disc, or an eliptic shape that on one side goes apart where you have the power in the context of negative states for the anger and fear. In any case, this is quite complicated. However, if you want to model emotions, you will have to deal with the fact that there are these two different ways of looking at these things. And that physiology might map much easier at the level of a few dimensions, but people’s understanding or naïve representation is much more linked to discrete states. So if I try to explain this to you, it is very abstract. But you have thought about these things. If I was to grab somebody from the street and would try to explain this dimension thing, it wouldn’t get very far. If I ask them what are emotions they would say it is things like happiness, love, and whatnot. So basically, that might be important at some point. And maybe if you are looking for names, somebody who is important today, apart from Peter Lang whom I mentioned, and Margaret Bradley, that would be James Russell in Boston, who has been talking about dimensional approaches for the last 20 or 30 years basically.

Kappas – I will now rewind to the very beginning where you asked me the first question. About categories and basic emotions and things like that. So the point is that for somebody who belongs to these dimensional approaches, that question does not make sense.

Seaman – You did talk about heightened probabilities…

Kappas – That was a dimensional approach also.

Kappas – When we were in the third type of emotional theories, the cognitive theories, we talked abut Scherer, who was an “appraisal” theorist, and cognitive theorist, and for him everything can be explained in terms of dimensions, however it is not a 3 dimensional world but the n is much higher. For him in this high dimensional world, emotional states are dots in that space, that move through there as you change. That there are particular links between bodily responses, and where you are in the space. And that some of these places in that space could be seen as points of higher probability — her calls them modal emotions. He has sometimes made the reference that it is a little bit like strange attractors. Where things tend to come together but Scherer is also a dimensional person. Usually people would say he is an appraisal theorist. His webpage, has almost every paper that he wrote as PDFs. And so that is a very good resource. He is also very important — Klaus R. Scherer is a German born psychologist who did his PhD at Harvard, came back to Germany, developed in the early 80s an appraisal theory, was one of the founders of the International Society for Research on Emotion and about 10 years ago, with funding from the Swiss National Fonds, created what I think is the largest center world wide for research on emotions— a network of something like 10 universities, lots of money, loads of people, different disciplines— it is a world wide center for that stuff.

Seaman – And you were there, right?

Kappas – First I did my master’s with him, and later I did my Post Doc with him in Geneva

(we talked about Dartmouth yesterday) When I was done with Dartmouth I went to Geneva for three years to do a Post Doc working with him. At that time the Center did not exist. He later also was central in the creation of a network which is called “Humaine” and which relates to human / machine interaction and emotions. And basically that relates to what we now would call affective computing. “Humaine” still exists, but it is now an international society, that hosts conferences for Affective Computing and intelligent interfaces and is called Association for the Advancement of Affective Computing (AAAC). So Scherer is a good place to start. His work is quite complex. However he is getting more and more attention in different fields. So for example if you go in Europe to Media Psychology, people will talk about the effect of films and things of that sort, they will refer to his work. That is the short answer concerning the question of basic emotions.

What is the second question?

Seaman – How do emotions contribute to our sentience?

Kappas – Remind me what sentience is.

Seaman – Our understanding of the world… our general sense of being.

Kappas – OK. In terms of evolution – most likely, emotions pre-date consciousness. For some people, emotions are the cradle of consciousness. When I talk about simple organisms like the flatworm, a flatworm clearly (At least to some) shows emotional responses, but we do not know how much the flatworm is aware of the world. Or whether a flatworm is directed by what it wants.

Seaman – Hardwired?

Kappas – It is not only hardwired because they can learn. So you can teach them things. As soon as you can learn.

The idea would be that emotions are a guiding force. Without emotions it would be very difficult to set goals for us. So the idea is that we have certain needs— you need to eat; you need to drink; you need to go to the toilet; these things are pretty basic. And it is not that there is simply a flag that is waved, but hunger doesn’t feel good. And people pack food because they are afraid of being hungry or afraid of being thirsty. Quenching your thirst is a very pleasurable thing. So what you find is that motivations and emotion have a very tight dance with each other. Which is one reason why for example my course following long traditions is called 'emotion and motivation'. Or there is a journal called Motivation and Emotion. Because these two things are really closely related. Emotions have motivating forces. And motivations work by using emotions if you will. So, if you start from scratch, and you would build something. One of the things that would happen is that thing would need energy. So you need to find a way, so if the thing doesn’t get energy, it goes and finds energy. Otherwise it will die. So this means you would have to have a very basic mechanism that makes it desirable, that sets a goal, to get the energy if the energy goes down. It might be the less energy you have the more urgent it becomes. So you need a system that is able to account for it. And one way that the motivations would talk to you is via subjective experience of emotions. People like Damasio have argued in the books already mentioned Descartes Error or The Feeling of What Happens, that it was a huge mistake to separate reasoning from emotion. And make it appear as if logical reasoning, rational reasoning is what we should aim for and emotions are just distracting from it. He makes the point that we need the emotions to make good judgments, good decisions. So there are patients that have difficulty reading their own emotions, in that respect. And when you want to make an appointment with such a patient you never get to the point…. well I could on Thursday, well there is Friday, well there is also Monday. There is nothing that helps them to jump if you will. In everyday life most of the decisions that we make involve too many things to ponder to think them through. If you have a computer that plays chess. The simple way of doing it is you take an infinitely complex computer, that simply computes all possible moves. That is not possible for practical reasons.

Seaman – The brute force approach.

Kappas – Yes, brute force. You need short cuts. You need heuristics. Or intuition. And intuition and emotions are closely linked. So when you are in a city that you do not know very well. You need to go to the train station. You are a bit worried that you might miss your train. It is getting dark and there is a narrow alleyway. It could be a shortcut. You do not need to think — that could lead somewhere else. There might be strange people in there. It might be dangerous. Instead it just doesn’t look good. It doesn’t look desirable. You don’t want that. And so the idea would be that the emotion, the subjective experience of the emotion is like a message system to our consciousness that alerts us to a relatively complex state of affairs. So rather than having to ponder all of those things— you like this, you are afraid of that, you are angry against that person, and so in that sense the emotions contribute to the sentience by providing guidance, by reflecting very complex states of person/environment transactions and at the same time as part of that function it is not just a messaging system, but at the same time it prepares the body for potential responses to the situation. So one of the old notions is the flight /fight response. So not only does the organism get the message that you need to act now, but various parts of the body go into high gear facilitating running away or attacking. And now we have a much better appreciation for how deep that goes. For example even including the immune system which is modulated by these things which has an impact on how well things heal… Your digestion is being turned on and off, things like the sweat on the palm of your hand that we measure in the laboratory has to do with you being able to grab things - it gets sticky which is good if you have to go up a tree. So it is both a messaging system as well as something that distributes the resources in the body. Both resources in the brain, by directing attention, by facilitating access to memories, by influencing the categorization processes, but also in the supporting systems such as the cardio vascular system and as mentioned the immune system. So the influence of emotions on sentience is cannot be overestimated.

Seaman – So here comes a really tricky one… So we have these feelings in our body that motivate us, how would those feelings be embodied in a machinic system?

Kappas – In the Cornelius book that I mentioned yesterday as the Intro Emotion book, he makes reference to a science fiction that I read too and I am trying to remember the author (Anne McCaffrey), I think the name of the thing is 'The Ship Who Sang'. And the idea was that you had these people who through accident or whatever loose their body. So they keep the brain alive for example in this case and implant it into a space ship. And so that human brain, can sense all of the mechanistic aspects of the ship as if it was a body, because it becomes then a body for that brain. The story has to do with this idea — how would it be if your brain were to be put into a mechanical body.

Seaman – Sounds like the Hillary Putnam article – Brain in a Vat

Kappas – Yes, yes that is a classic. Now this has been around as a gedanken experiment for a long time but we are getting into the more concrete thing because not only do we now start to give people who have lost control over parts of their body, robotic arms or things like that. But it is less unusual these days that you have somebody doing remote control over a robot for example, and they start to feel as if the arms for example become their arms. So there has been all kinds of attempts for providing tactile feedback, however, we get a sense of tactile feedback even apparently based on things like visual feedback, because of the multimodality of our perceptual system. I think I might have talked the other day about the rubber hand illusion.

Seaman – I think so.

Kappas – The rubber hand illusion where you put a rubber hand in front of somebody, and then you block the vision so that you cannot see your own (let’s say) left hand, because there is a wooden board blocking the view but instead you see a right hand and you see that rubber hand that is oriented almost as if it were your hand but it is disconnected from you. And then somebody takes two feathers and at the same time strokes your hand that you cannot see and the rubber hand. And after a while, when you look at that rubber hand you feel that this is where the thing comes from. So that when you then approach with the knife, the rubber hand, people freak out.

The same thing might happen with robots. So that simply because our nervous system has such great plasticity, we start to feel even artificial limbs, as if they were our limbs. I just had a whole bunch of crowns done (my dental work) and when I had the temporary things it felt like I had lego pieces in there that had nothing to do with me. Now I have the final things and they feel like teeth to me. It is like — this is a part of me when I eat. I don’t feel that there is a “plastic” thing or a metal thing between --- it is part of me. So in that sense we do know that human brains probably can get feedback from artificial systems. The question is how would an artificial brain get feedback from artificial systems? And would it be an emergent property or would we have to design it in.

Seaman – Or a mix maybe?

Kappas – Or a mix.

Kappas – Well we do know that somebody who does not experience pain is at risk of dying very quickly because they bump into things, they don’t realize that they do damage to their body, things get damaged quickly. So if you were to design an artificial system, that could move about, you would want to have such a thing as pain. And then it becomes simply a philosophical thing — to what degree a read out that drives an if/then and/or/else loop is the same thing as you having this subjective experience. And that is a separate issue.

Kappas – I would assume that you could mimic some of the advantages of emotional systems without trying to tackle the problem of consciousness. But you would want to have motivating systems in there, that use some sort of emotions simply because resources and time are limited and these are highly efficient ways of managing resources, and managing tasks, under conditions of constraint.

Seaman – It is interesting to think about because adding machinic senses to us, we have in our brains this potential for feeling whereas if you ask what is that potential exactly, and can you make a machinic version of it— it is a very abstract kind of thing to think about what feeling is.

Kappas – I remember many years ago… quite a few years ago I was in Canada. I was at a conference of cognitive and biological (whatever) society. And there was the reception and we were sitting at round tables. At the table where I was were several people who were interested in faces. We knew each other. And then there were these few people that none of us knew and we said – so what do you do? And they said – oh we work with ‘water runners’. These insects that can run on the water. And so we all looked very puzzled. Why? And Why at such a society? Well, he said, the nervous system of these animals has like 10,000 neurons (I don’t remember the exact number he said). We can model the complete network. And that just hit me. What that means. So when we talk about modeling emotions, any model that we build will be so simplistic with regard to the complexity of our brain. Of course we can not approach anything like the type of consciousness that we don’t have from the outset. But that we have to learn over years basically. However, if I would say, would I be able to model a lizard? A lizard is much more complex than one of those insects.

Seaman – There is a fish that Rössler is interested in modeling – the Pandaka Pygmaea. He wants to create a foundation to do just that.

Kappas – So that is possible. Quite in reach. You will say – will that fish have some emotional systems, in there, then yes. How do you know what the fish has for emotions. Now if you grab the fish and the fish is trying to escape there are some emotional implications. The other day I read an advertisement and it was Paul McCartney and how he became a vegetarian. And he said that he was fishing and he was taking the fish off of the hook and the fish was bouncing and flopping around. For him at that moment he had the feeling this fish wants to live just as much as you do. That was for him a moment that was very powerful. He put the fish back and he stopped eating meat. Because from that moment on, this was how he felt about that. So if you would say looking at a flopping fish in that situation, that there is something emotional to it, it will be very easy to have the artificial fish doing the same thing. Without having to spell it out. That will be an emergent property of certain things that you put in. If we call that a shallow level of emotionality, then yes I believe we can do that. This is within our grasp. I do believe in all kinds of rapid advances. But I am not convinced that in our lifetime we will see artificial systems with the complexity of a human. But again, I really don’t need that to grant an organism emotions. At that point we get into ethical issues.

Seaman – It is also about this issue of the subjective report. I can’t feel your emotion and yet I understand that you have emotions because you told me you have emotions. So it is a very interesting case that when you have a robot that you are creating, and it might tell you, ‘yes, I feel emotions’.

Kappas – Yes – and what would you do? What would you do if Hal says don’t switch me off.

Seaman – Which he basically does.

Kappas – Which he basically does. Do you believe that?

Seaman – This goes back to Descartes, to his meditations. This dream…

Kappas – The thing is our subjective experience of interacting with these entities is a completely different story because we get sucked in so easily. As children we can take dolls and put everything into them, it doesn’t take much.

Seaman – Cynthia Brazeal at MIT did just that --- there is no mind there. Yet, people are talking for ½ and hour to this thing.

Kappas – They get sucked into it. And for one of my courses – Communication and Interaction, I use an ‘Eliza’ version. So every year when it gets to that point I go online to see whether I find a new ‘Eliza’. People are constantly working on these bots.

Seaman – In fact artist Lynn Hershman-Leeson is one of the people.

Kappas – my colleagues in CYBEREMOTIONS at the OFAI [[Austrian Research Institute for Artificial Intelligence]] – the group that Robert Trappl heads… they have these things.

Kappas – So I am trying this out, and it is pretty much obnoxious and I want to quit and I say “bye”, and it says "don’t leave now" and I write “I have to go”, and it says “you didn’t like it did you?” “No” “I have to work” and it replies “Just be a bit longer with me” or something like that. And I know how it works. But still it drives you… it drives you nuts! It amuses you.

Seaman – It is almost uncanny.

Kappas – So the question… I think it will be much easier to create systems that feel like they have emotions. That feel like they understand us. Where we feel empathy… but if we were to look at it from outside of that immediate experience as observers looking on we would not be convinced that that actually is what it is.

So depending on what the goal is that you have in building an artificial entity, you always have to go for what it is that you want.

Seaman – Well if emotions are part of intellect, and driving us, then it makes sense in the construction of this highest level intellect system to have them be part of it, as a kind of logical argument.

Kappas – The thing is that… What you do not need to do is to build an emotion feeling service robot ... something that sells you your jeans, Or that is an agent that is not physically embodied working for the amazon on-line shop. You could imagine doing something such that if you ask for something they don’t have the thing looks sorry. And you really have the impression – that it is sorry. Or are annoyed you had to wait for a plane, and you see the thing on the other side of the screen blushing, looking away, swallowing hard, and you might stop yelling at the system because you feel pity for it. I think that is easy to do. But in this case the thing need not really be depressed.

Kappas – If what you want to do is to build a human-like entity, then of course it must be part of it.

Seaman – Arbib wrote a book called Who Needs Emotions Anyway. There are some people that think this is a very big mistake to do this.

Kappas – Yes, of course.

Seaman – But it is also a moment of complete social change, if a robot has this kind of notion of personhood, true emotions and so on, it no longer can be a slave in a sense.

Kappas – Yes, and this is something that Isaac Asimov has written extensively about in his Robot novels. And one of my favorite Star Trek episodes that I have used in class occasionally, it is Star Trek Next Generation and they take away 'Data' (the robot) in The Measure of a Man and they want to disassemble him. And Data says he doesn’t want to. And then there is a court hearing whether they have the right to switch him off and disassemble him or not. Despite the fact that this is mainstream television, I thought they do get the point across very well. And in the court hearing, Riker who defends Data, does something, he switches him off, and this is incredibly powerful because we feel for Data. And he switches him off and shows that ‘it is a thing’. And what we know as Data is gone at that moment. And then he switches him on again to let you reflect on what that actually means.

And that is very very brilliant.

Seaman – Because of death. Because we can be switched off. But we can’t be switched back on.

Kappas – So part of the strong reactions that people have about putting emotions into things has to do, no only with out own fears in regard to enslaving machines, but our own mortality, the definition of who we are, and I would assume part of your Neosentience project is has to do with the realization --- who are we?

Seaman – Yes. There is also this whole techno-dystopian history --- the fear that these robots will take over and be smarter than we are.

Kappas – As I say in the course sometimes. I do believe in the “topical T-Shirt”. So I am wearing a topical t-shirt today. Do you know what that is [[what was on the shirt --- I can’t remember]]

Seaman – Gave a whole paper on this topic. It goes way way back— the Golem, then we have Mary Shelly.

Kappas – Actually yesterday evening I had an interview with somebody from the New York Times who was writing a paper on Affective Computing. And we got to talk about the Golem, and mary Shelly and all of that. It goes back and probably there are myths in less technological societies that we don’t even know, that relate to that.

Kappas – This notion of putting life into a dead thing and what does that mean, is really a very powerful story line that we can all relate to. Despite the fact that as children, that is what we do! All of our puppets take on their own life. And this is normal. Many of the naturalistic religions put life into objects.

Seaman – The Animism of the Japanese for example.

Kappas – However it is not us who puts it in there, they are in there, they are part of it… The difference is of course is that here the human becomes the creator and this clearly means that the human tries to play god, by creating himself at that moment. And so at that moment it touches on many many taboos. And it is funny because… it is clear that there will be religious taboos, but how is it if you are perfectly secular being, and you are not religious, and you are not spiritual at all, would you still be offended by the idea. My guess is many people would because it's a very fundamental thing to do.

Seaman – Yes – I have had people in class say ”Why would you ever want to do that”?

Kappas – There I would distinguish between trying to understand what the motivation is as opposed to judging the motivation of thinking that it is false or misguided.

Seaman – I think it is a mix.

Kappas – It is a mix – yes.

Seaman - I think this is a natural ending point.

Kappas – Yes I agree.

Seaman – Thanks Arvid for taking the time to discuss this with me! It was very helpful.

May 19, 2013