The path towards the integration of several technologies to the autonomous systems of the near future is covered by several challenges. Some of those challenges are described in this short paper. The format of the paper is in the form of a dialogue between two discussants, Socrates and Melambus. For reasons of brevity, no prerequisites or references are given.

1. Inherent bias

M: Socrates, there is immense interest in the Science of Autonomy. Many groups are working on the problem: lots of meetings and workshops, papers, journals, funding,.

S: Which fields are involved in Autonomy?

M: There is a cornucopia of fields represented: there are people from Control Theory, Robotics, Computer Vision, Computer Science / Programming Languages, Machine Learning, Cybersecurity and Artificial Intelligence. Most of the research is restricted to one discipline but a few investigators embrace more than one discipline.

S: You already see the first difficulty in designing autonomous systems. If you ask a control theorist what is the most important component of autonomous system, he/she will tell you: “obviously, the control system. Without it you cannot guide your system – period”. If you ask a perception expert the same question, you will probably get the answer: “Do you think it is random that about half of the brain is doing vision? Without perception you know nothing about the changing world, so there is nothing to do – without perception you cannot have an autonomous system – period”. Similarly, experts in other disciplines magnify what they understand best.
M: Hmm – I see. You are talking about an inherent bias on the part of the scientist. But what are the consequences of this?

S: If you now ask the researchers to design an architecture for the autonomous system, the control guy will design a control system and will glue the perception and the rest of the components appropriately. The perception guy will design an active perception architecture and will glue the other components, and so it goes. But why this sudden interest in autonomy?

2. Self-driving cars

M: Well, it’s a challenging problem but I think the interest is amplified because of the self-driving car industry – you see, a self-driving car is an autonomous agent.

S: Why self-driving cars? Is there a need right now for them?

M: Well, was there a need for cell phones? It is progress, but there is a very strong argument in support of this. It goes like this: 90 percent of the car accidents are due to human error. So, if we take out the human then we will save that 90 percent and this translates into thousands of human lives.

S: Wait a moment, this is a faulty argument. You left out some things. You will take out the human but then you will replace him or her with a computer. Right?

M: Right.

S: And on top of this you assume silently that the computer will not make any mistake. Right?

M: That’s right, but you know that’s what computers are for – they don’t make mistakes…

S: No Melambus, this is a wrong argument, rather an incomplete one. I would assume that there exist elaborate simulation systems that can simulate thousands of driver less cars moving in a city and develop an
estimate of what happens to the 90 percent. For all we know, maybe it becomes 95 percent.

M: Unfortunately, as far as I know, there are no complex simulation procedures that would allow testing under a very large variety of conditions. It is a PhD dissertation topic in a few prominent universities…

S: It’s worrisome indeed, but hey, that’s capitalism, right? In any case, yes, the self-driving car is an autonomous agent.

M: And so the problem becomes: what is the architecture of the whole thing?

S: You have of course the same problem for any autonomous system. You should focus on the task and this will tell you which aspects of the environment you should take into consideration. These will become your variables in the modelling of the problem. But before we go on, let me tell you one last thing about the driverless cars. When you replace the human with a computer, the assumption is that the computer is at least as good as the human and hopefully better. This is of course with regard to driving. The human “sees” things out there and adjusts the controls with his hands and feet. The human, besides his visual capabilities that we try to replicate, he also has common sense reasoning abilities. Unfortunately, we don’t know how to solve this problem yet, i.e. to give machines common sense reasoning. We have theories and potential explorations, but no solution yet. The industry will take longer to materialize than is commonly promised.

M: Socrates, for heaven’s sake! What do you need common sense reasoning in order to drive?

S: You need it for the situations for which your system was not trained on. You can’t train for everything! You train for some things and then common sense tells how to deal with contingencies.

M: OK, you have a point but I am mostly interested in an integration methodology.
3. Autonomy = Integration

S: Indeed Melambus, the problem of autonomy is the problem of integration.

M: The problem is that usually “No credit is given for integration work”. I wonder why?

S: This is true indeed. The reason, perhaps, has to do with the fact that most integration work is an afterthought.

M: How so?

S: Well, take Joe and Jim and Ned. Joe is an expert in Computer Vision, Jim in control theory and Ned in planning. And then Jim says: guys, if we take Joe’s stuff and Ned’s stuff and we integrate it with my stuff, then we have a system! You start from the fields – this is a given, since all these different communities (computer vision, planning, control, etc.) have developed many useful tools. And then, you will have to put them together. There is a methodology for doing that. You design the modules of the system and you start testing them in pairs, in triples and .. all together.

M: For example, if I have to do autonomous landing using vision, then I would need to extract the appropriate representations of the terrain using computer vision and pass them on to the controller in real time. The problem can get too involved because depending on the output that you choose to consider, the complexity may be changing drastically.

4. Bottom-up integration

S: Good! This is an example where you will have to integrate computer vision with control to close the loop. If you have a different task, you will integrate differently. Or, if the planner needs to be integrated with perception, that integration will also be different. So, think now, all these different tasks and all these different processes (vision processes, planning
processes, control processes, and so on) give rise to all these different integration tricks. This is perhaps why no credit is given for integration: most integration resembles some form of alchemy, where people try out various possibilities. If you think of the whole system that integrates perception, with control, planning and reasoning then, assuming that many tasks are possible, the integration resembles a structure like the biblical “tower of Babel”.

M: Ugly or not, that integration will work and we will be done!

S: True, but we still don’t have a methodology. We haven’t figured out the periodic table yet, we are still at the alchemy stage.

M: What do you mean?

S: Look – let’s say you integrate the different processes in your autonomous car. Is that the only autonomous agent around?

M: No, we will have many autonomous agents, on the road, in the air and on the sea.

S: OK then, for each one of them you will have to integrate again for the specific task. And you will have to deal with each case from scratch – as if you didn’t learn much from setting up the integration of the driverless car. And let’s not forget that the car is conceptually an easier case because it has restricted movement (on a plane). Imagine a drone whose movement is unrestricted rigid motion, i.e. instantaneously the sum of a rotation and a translation. This is a much harder case.

M: I understand your point, but isn’t this engineering?

S: Yes, it is if you adopt the viewpoint I just described to you. You take a little bit of this with a little bit of that, you integrate them for a task, and you keep going. It is a bottom up approach.

5. Top-down integration

M: It doesn’t seem there is another way.
S: There is always another way – in this case the top down way. Adopting it also allows you to cast the autonomy problem as a scientific problem.

M: How so?

S: Look around you Melambus – you see the bees, the birds, the squirrels, the humans. They are all autonomous systems! How are they structured? How do they work? This is a good question in the Sciences and lots of people from many disciplines are studying aspects of it.

M: But Socrates, we don’t really know how all those systems work!

S: That is true, but we know enough so that the process of “inspired imagination” can take over. Think Melambus, what do these autonomous systems, the bee, the bird, the squirrel, the dog, the human have in common?

M: What could I, a human, have in common with an insect? Socrates, I don’t get it.

S: Would you say that bee knows what the beehive is? I mean, does the bee know its home?

M: I am sure it does.

S: Would you say then that the bee recognizes its home and knows what to do with it? In other words, it has more knowledge about the hive, how it is inside, ways of moving, and so on.

M: Certainly. I would say that bee knows even more things, about flowers, distances and they also communicate that information with a dance.

S: Exactly! Would you say then that the bee has a number of concepts?

M: Concepts? Like humans have concepts? I thought concepts were a human quality.

S: Humans have very elaborate concepts, but no one prevents the bee from having some concepts as well. They will not be elaborate, they will be simple, but they will be concepts.
6. Autonomous systems contain conceptual systems

M: What are concepts really?

S: They are just knowledge along with relationships among those pieces of knowledge. So, the bee has knowledge of the hive not only in recognizing it visually but also it has some form of a spatial model, much like you have a model of your house.

M: Socrates, if you tell engineers about concepts, they turn the other way. It becomes cognitive, too high level for a mechanical engineer. Is it necessary?

S: Of course it is necessary. The argument is that since these systems are autonomous, they have an understanding of the events around them. And you can’t understand events without concepts. Take as an example your favorite driverless cars. Do they have the concept of a car?

M: I guess so. They should, they recognize other cars under any conditions.

S: Good. But now that they have the concept of the car, they have more, because the car has four wheels, for example, and we know that.

M: I see. When you are at the concept, you bring all the knowledge that you have about that concept.

S: Exactly. And this allows you to do better reasoning. Let’s go on. Your driverless car has more concepts, right? Like “human” or “animal”.

M: Of course.

S: Now events are particular actions relating the concepts. Like “a human is walking in front of my car”.

M: I see. Your argument is then that what is common in all those autonomous systems, the bee, the bird, the human and the driverless car is that they all have a conceptual system.
S: Exactly! And that’s the thing to start from. The AI people have lots of ideas how to deal with concepts.

M: But how do you know what concepts to start with? And how do you proceed, you give them to the system or you learn them?

7. A methodology for integration through tasks

S: Ahh, Melambus you hit the million dollar question, as they say. Let’s delay answering it. Instead, let’s ask a pragmatic question: assuming that you want to design an autonomous agent, you should know what this agent should do, in other words you should have a set of tasks $T=\{T_1, T_2, \ldots, T_n\}$ that the agent should autonomously perform. In order to perform those tasks, the agent will need to have a number of concepts $C=\{C_1, C_2, C_k\}$. These are the concepts to begin with.

M: Hmm, so then given the set of concepts $C$ the problem of autonomy becomes the problem of integrating the perception, the planning and control with the concepts!

S: Very well said Melambus. Of course all these autonomous systems will be sophisticated and intelligent expert systems, but you have a chance to study the problem of autonomy and intelligence this way. You can think of systems that continuously learn, like humans.

8. Primitive concepts

M: But how do you start then? What are the concepts you start with?

S: The fields of Neuroscience for quite some time they have argued for innate concepts. You are born with a number of concepts and then you build new ones on top of the innate ones, and so on. Think of those innate concepts as primitives in your system.

M: But what are they?

S: We don’t know exactly what they are, but we have some good paths to explore. After all, with regard to autonomous intelligent system design it doesn’t much matter exactly what the primitives are. What matters, is
whether you can combinatorially combine the primitives to create unlimited situations.

M: Can you give some examples of primitive concepts?

S: Touch. Whether you touch or are touched it is a sensation directly obtained from the skin. You actually know what it means to touch or be touched before you learn the word: “touch”. The same is true with left and right and the spatial prepositions. Try to define “left” in some general way, without referring to a specific scene. You can’t. The physicists know that, that’s why they say “clockwise” for example, which assumes the knowledge of the concept of the clock. When the children learn the word “left”, they already know what it means. There is a set of such concepts whose understanding comes directly from the sensorimotor information.

M: I see. I can now start understanding the world by putting those concepts together. I can recognize events and I can learn new concepts, by combining what I already know. I can’t wait to tell my driverless cars buddies – you start with the concepts and you integrate by getting every sensory-motoric signal to the concepts. Then knowing the events around you, you plan your autonomous task.

S: Well said, but it’s not so simple. You have two additional problems you need to guard against.

M: What are they?

9. Active vs passive vision

S: The first is that the bulk of contemporary computer vision is passive and disembodied, while the vision of autonomous systems is active and embodied.

M: Active, passive, what is the difference. Vision is vision.

S: There is a big difference Melambus. Like night and day.

M: Why. Can you give me an example?
S: Take tracking as an example. In the passive approach, you are given a video where something is moving and you want to put a rectangle around the moving thing, and follow it through its image movements. In the active approach, the one that your autonomous agent will need, you have a camera on a motor. The camera is looking at a moving thing and now the motor has to move the camera so that the moving object is always at the center, much like you do when you track things by moving your eyes. In the second case, you get new information as you do the tracking, something that does not happen in the first case. The two cases amount to different problems.

M: But Socrates, what you call passive vision has given breakthroughs in the recognition problem. Now we have deep learning networks with performance in the 90s. Nine times out of ten they are right.

S: Would you enter an airplane if it was 90 percent right? It would have to be 99.9999999.....99 and the question is how many 9's! Melambus, you miss the connection between recognition and the real problems of search engine companies and social networking companies that dominate the field of computer vision. This algorithm that has 90 percent performance in recognition is not adequate for the autonomous agents that need to recognize. It is a breakthrough for a search engine, because if you ask the engine for images containing “blah”, nine out of the ten results that the engine will return on the first page will be correct. Indeed a breakthrough. But you cannot give the same algorithm to a robot to recognize “blah”. You have to do more work. Not to mention that most recognition theories are based on single images and autonomous agents do not perceive single images. They always get a video. So, the recognition breakthroughs of Computer Vision do not transfer to Robotics. For this, Vision needs to be Purposive, Selective and Directed. The real problem in vision is what to keep from the image and what to throw away.

10. Embodied Perception

M: You also talked about some embodiment. What is the big deal about embodiment.
S: An embodied system is one where its motor representations are indexed together with the perceptual representations. For example, I know how far you are from me because I know how I have to move my hand to touch you.

M: That’s it? That is the whole point of embodiment?

S: There is a consequence of this that makes embodiment essential.

M: What is it?

S: It has to do with prediction. Every system that moves in the world, has the potential to relate its “motor” sequences with the “visual” sequences.

M: OK. But systems move differently; some of them crawl, some fly, some jump, walk, and so on.

S: Precisely! But when you fix the system physiology and mobility characteristics, you are in business, because you can now index perceptual stuff on the back of motor stuff, which is much simpler. In other words, you can learn using modern techniques to predict what you are going to “see” next.

M: Hmm, but that way if I can predict what I will see, what is perception really?

S: It is some form of a controlled hallucination process. You hallucinate the model and you keep checking if it “fits” the incoming visual stream.

M: And what if something unexpected happens?

S: These are the contingencies – they break the symmetry and now you have to deal with them using common sense reasoning.

M: This was the first thing regarding obstacles in the development of autonomy. What is the second.

S: The second has to do with how the autonomy scientists perceive perception. How does perception work?
M: We have the theory of David Marr for this. Perception provides a world model – it is a global 3D model that is then given to the Planning processes to plan how to achieve a task.

S: That’s right. This has been very useful framework as the field was developing, but there is an alternative viewpoint on the nature of perception, which is closer to reality.

M: What would that be?

S: It originated in the work of von Helmholtz, the famous physicist. He was also interested in Perception. Towards the end of his career, he introduced Perception as Unconscious Inference. He argued that images alone are not enough to create an understanding of the scene in view. So, he reasoned, as we look at the world, we also think about it, but we are not aware of this thinking and thus he called it unconscious inference.

M: So, in some sense perception and thinking are having some back and forth, with perception influencing thinking and thinking influencing the nature of perception.

S: Exactly – perception and planning are two sides of the same coin, with the process being controlled by what is known as attention.

M: Yes, there is too much written on attention. What is attention anyway!

S: Attention is really the operating system of your autonomous agent. It regulates how you deploy your limited resources to problem solving.

M: Yes, but we have powerful computers. Couldn’t we work on the whole image anyway?

S: There are too many events happening at the same time and there is a lot of uncertainty. Let me give you an example. Let us consider the following situation from a urban scenario where a swarm of autonomous robotic systems is required to perform a number of tasks. These robots may be required to explore and make repairs to a scene, find people and communicate information to a command and control center. They must
perhaps locate leaks in some tubing or they must find a special control box, while at the same time watching for humans and listening for “cries for help”. The robots must make sense of the cluttered audio-visual environment. The audio environment is cluttered – with alarms, hissing sounds from leaks, gunshots and self-noises from the robot’s own motion; so is the visual environment, which after a number of explosions only bears a slim resemblance to the visual plans the robots have in memory. How can the swarm find task-relevant objects in this situation?

M: Hmm – that’s a challenging situation indeed, much harder than self-driving cars.

S: In this situation the robots are required to identify – find, locate, segment – an object. It could be a visual object, an object we can see. It could be a small object we can grasp (like a weapon or a valve), or a large object (like a table or a house). It could be a human body part (like a face, an arm or a hand), or a body part together with an object or a tool (like a face wearing a special hat, a hand carrying a pistol or a screwdriver). It could be an action itself, like a human pressing a button or a human digging. It could also be an auditory object, a sound, mixed with the multitude of competing environmental sounds, that is associated with a particular object or an action (a voice, water running, a glass breaking, a car passing by) like the ones just described.

M: I see, but why couldn’t I develop deep learning for all these “objects”?

S: Because you will have a lot of uncertainty it’s hard to do it the classical way. Remember, according to Helmholz, you need to think at the same time. Imagine you enter your kitchen, after a party, to look for a particular pair of scissors. What would be your search strategy? Would you try to remember where you last saw the scissors? Or would you try to go for the obvious locations of where scissors would be placed – in the drawers, or besides the knives? Once you have prioritized where to start searching, you start to remember how your particular pair of scissors looks – its shape, size and maybe some unique identifying color or labels that could enable you to discriminate it from other pairs of scissors that have other uses. By mixing signals with prior knowledge, some of which is in symbolic form, humans are able to cope with this kind of open-ended problem. To achieve such a remarkable solution, they utilize a very elaborate “attention
system” that guides them to find the “next object” that is critical for accomplishing a task (one that may change dynamically).

M: Hmm, I am still not getting it.

S: Let us examine in more detail what is going on when a human (a cognitive system with vision and knowledge) is interpreting a visual scene. When we fixate at an object, attend to it and recognize it, it results in an immediate entry to the conceptual system. Indeed, if we recognize a “street”, the concept street “lights up” in the conceptual system, with a number of consequences. The word “street” has many “friends”. These are other concepts that tend to co-occur with street”, such as “human”, “car”, “house”, etc. Thus, recognizing a concept in the scene creates expectations for the existence of other concepts in the scene for which vision can check.

M: Ahh, I see. As you interpret a visual scene, you fixate at some location and recognize concepts (nouns, verbs, adjectives, adverbs, and prepositions). Because the conceptual system is highly structured, these recognitions produce a large number of inferences about what could be happening in the scene. This leads you to fixate at a new location, and the same process repeats.

S: Exactly. This kind of structured exploration guides the auditory system as well. Because every sound is the result of some action, recognition of a sound leads to recognition of a concept (verb), which creates expectations for the involved objects and tools, and so on. Thus, an interesting new way to study attention is to integrate the senses with the intellect, linking vision with knowledge.

M: Thank you Socrates, it has been a pleasure.

S: Good luck! And remember Marcel Proust: The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.