Goals emerge in macroscopic descriptions of the world[1]

Stefan Keppeler

Fachbereich Mathematik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
stefan.keppeler@uni-tuebingen.de

Descriptions of the macroscopic world typically feature goal-oriented dynamics. A virus attaches itself to a host cell *in order to* subsequently make it produce more viruses. A plant may grow leaves in a way *such that* it can collect more sunlight for photosynthesis. You may decide to continue reading this essay *in order to* learn about my views on goal-oriented dynamics. In these examples we explain the behavior of macroscopic entities not in terms of the their present state but in terms of possible results of their behavior. We perceive the virus, the plant or you as agents pursuing goals. Macroscopic entities of this kind are no longer mere objects whose dynamics follows a predetermined track but instead we perceive them as, at least partially, causing their own dynamics.

Macroscopic entities, in the context of this essay, may be as large as humans and other mammals or as small as bacteria and viruses. I reserve the word microscopic for the realm of atoms and molecules or even smaller building blocks of modern physical theories. Even though, e.g., bacteria are microscopic in the everyday sense of the word I will refer to them as macroscopic entities since they are much larger than, say, atoms or molecules and since they consist of a large number of the these microscopic building blocks. Theories describing the latter are formulated in the language of mathematics. Goals, however, are absent from microscopic theories. No matter whether we describe the world in terms of classical or quantum physics, no matter whether our theories are deterministic or probabilistic, goals have no role in these microscopic theories.

Since macroscopic objects are built from microscopic components this situation seems paradoxical. Are the microscopic theories incomplete? Are our microscopic and macroscopic theories of the world incompatible? Do we have to magically add anything new on macroscopic scales?

In this essay I will argue that goals emerge when we pass from a microscopic to a macroscopic description of the world. I will explore under which conditions goal-oriented behavior of macroscopic entities can emerge from goal-free microscopic dynamics. I will show that this emergence is no mystical process in which the microscopic laws of nature suddenly cease to be valid when a certain number of atoms or molecules is gathered in one place. Instead I will explain, that macroscopic theories are necessary in order to describe the world on a larger scale, and that these theories, even when they display qualitatively new features, are fully compatible with microscopic descriptions from which the very same features are absent.

A well-studied example of a feature present in macroscopic but not microscopic theories is irreversibility. Since the work of Boltzmann it is well-understood that macroscopic irreversibility is compatible with reversible microscopic laws. Before coming back to goal-oriented behavior I will therefore briefly review Boltzmann's explanation of irreversibility. Then I will argue that an understanding of goal-oriented behavior should be pursued along similar lines.

---

[1]Contribution to the FQXi essay contest *Wandering towards a goal: How can mindless mathematical laws give rise to aims and intention?* `http://fqxi.org/community/essay/rules`.

Irreversibility describes the fact that many, essentially all processes which we observe in the macroscopic world have a preferred direction of time. We may observe a cup shattering into many pieces, but we never see the shards assembling themselves into a cup. Or imagine me filming anything happening around me and playing back the movie for you forwards or backwards. You will be able to tell within seconds whether I chose the wrong time-direction. As common as this observation is it may appear paradoxical when examining more closely the physical laws which ultimately govern the observed processes. These laws, think, e.g., of the laws of classical mechanics, exhibit no preferred direction of time: If a certain sequence of events is a solution to the equations of motion, then the reversed sequence is also a solution to these equations of motion. In the case of classical mechanics consider a system of particles with given initial positions and velocities evolving for a certain period of time. Record the final positions and velocities of all particles, use these positions and the reversed velocities as new initial conditions, and let the system once more evolve according to the same equations of motion. Then all particles will move along the same curves as before but in opposite direction: The movie played backwards shows an equally valid solution to the equations of motion! This property is known as time reversal symmetry and it is no distinctive feature of classical mechanics but is shared by all microscopic theories.[2]

How can macroscopic dynamics be irreversible when microscopic theories are time reversal symmetric? Boltzmann's explanation is as striking as simple. There are many good accounts of Boltzmann's derivation of irreversibility. I particularly like Lebowitz' summary.[3] Macroscopic systems displaying irreversible dynamics are composed of a large number of microscopic components. We describe the state of the system as a whole, its so-called macrostate, in terms of a much smaller number of variables. For instance, some liters of gas in a container are conveniently described in terms of the density (in six-dimensional space of position and velocity) measured with macroscopic resolution. These few liters of gas consist of about 10^{23} particles (atoms or molecules). The so-called microstate of this system is described by the positions and velocities (and if applicable angular momenta) of each of these particles. Each macrostate can be realized by many different microstates. Imagine the space of all microstates. Different microstates corresponding to the same macrostate fill up regions in the space of microstates. Some macrostates, those which we observe frequently, correspond to larger regions than others, which we observe less frequently. I will loosely say that in the former case there are 'more' microstates corresponding to the same macrostate than in the latter case (although there are infinitely many microstates in both cases). In this sense, there are many more microstates corresponding to an approximately uniform distribution of the particles over the container than there are microstates for which all particles reside in the left half of the container. While the microstate evolves according to time reversal symmetric dynamics, when, say, the gas particles evolve according to the laws of classical mechanics, the macrostate of the system may also change from time to time. In most cases – for large systems essentially always – it changes from a macrostate corresponding to some region in the space of microstates to a macrostate

[2]For the theory of the weak nuclear force, which describes, e.g., beta decay of nuclei, time reversal has to be complemented by charge conjugation and parity.

[3]Joel Lebowitz *Boltzmann's Entropy and Time's Arrow* Physics Today **46** (September 1993) 32-38, `http://dx.doi.org/10.1063/1.881363`.

corresponding to a larger[4] region. Eventually, the system ends up in thermal equilibrium, i.e. in the macrostate corresponding to the largest number of microstates, and then it stays in this macrostate for almost all the time. In large systems, for all practical purposes, we can say that it stays in this macrostate forever. This is the irreversibility which we observe in the macroscopic world.

The striking thing about Boltzmann's argument is that it explains irreversible macroscopic dynamics with no need for microscopic irreversibility. Moreover, although I have illustrated Boltzmann's reasoning in the case of deterministic microscopic dynamics, his explanation works no matter whether the microscopic dynamics is deterministic or probabilistic, no matter whether our description is classical or quantum. Irreversibility is simply a consequence of our choice to describe a system consisting of a very large number of particles in terms of a much smaller number of macroscopic variables.

Here one might be tempted to ask whether irreversibility is only a feature of an effective theory, of an approximation, whether the world only seems to evolve irreversibly, whereas the real dynamics is reversible. I think this question is ill-posed. Of course, if we describe a large system in terms of each of its many microscopic components, then our description will be in terms of reversible dynamics. However, if we want to understand the behavior of this large system as a whole, if we want to describe how macroscopically observable quantities evolve, then our description, inevitably, will be in terms of irreversible dynamics. And this description will be empirically correct, qualitatively and quantitatively. On a qualitative level, just look around, and you will observe irreversible processes everywhere. And they are as real as the reversible processes which we observe in microscopic experiments. Put more dramatically, we are all born, age and eventually die – it never happens the other way round. What Boltzmann's argument shows us, is that irreversible macroscopic dynamics is compatible with reversible microscopic dynamics. And we may add that macroscopic irreversibility is not less real than microscopic reversibility.

Let me emphasize that I have not discussed Boltzmann's explanation for irreversibility because I believe that irreversibility was somehow necessary for understanding goal-oriented behavior. Instead I have chosen to discuss this example since it nicely illustrates how macroscopic dynamics can exhibit novel features which are absent from the underlying microscopic laws, and yet the macroscopic dynamics is derived from these microscopic laws. Could the explanation for emergent goal-oriented dynamics in a world which is, microscopically, governed by 'mindless mathematical laws' be similar to Boltzmann's explanation of irreversibility? Let us explore similarities and differences.

Irreversibility shows up in macroscopic theories like thermodynamics or fluid mechanics. These theories are formulated mathematically, say, in terms of differential equations, in the same way as the underlying microscopic theories. Consequently, these macroscopic theories allow for quantitative predictions which can be compared to experimental results. They are empirically equally well-established as their microscopic counterparts. Most theories of goal-oriented behavior are not formulated in terms of mathematical models. Are mathematical models for goal-oriented behavior conceivable? I think they are. Goals can obviously be

[4]Actually, it turns out that for macrostates corresponding to regions of different size, these sizes are dramatically different, and that macrostates corresponding to regions of equal size occur only under special conditions, e.g., due to symmetries.

incorporated in variational principles. The idea that an entity could itself initiate some of its own actions might be formulated in terms of inhomogeneous differential equations, i.e. equations including a source term which is not determined by the dynamical variables of the system. However, developing these or similar ideas into proper theories is beyond the scope of this essay – even for toy models. Thus, let us for the moment assume that for most examples of goal-oriented dynamics we have no well-established mathematical description at hand.

But without a mathematical formulation can we even refer to our descriptions as (scientific) theories of goal-oriented dynamics? I think we can. When we describe macroscopic entities as pursuing goals, we first observe the behavior of these entities, then model them as agents and make predictions about their future behavior. Finally, we compare their actual behavior to our predictions. We rely on these predictions in everyday life. We assume that the mosquito which landed on our arm intends to bite us and act accordingly. We even organize our whole society – political system, jurisdiction, economics – based on the assumption that the individual members of this society are agents, entities which initiate their own dynamics and pursue goals. Even when we do not have good mathematical theories for individual agents, we can quantitatively describe the behavior of large groups, e.g., a crowd of spectators or the stock market, in so-called agent-based models, i.e. by simulating these groups as composed of many interacting agents pursuing goals. Agent-based modeling[5] is a powerful tool, e.g., in economics and in social sciences.

I thus conclude that goal-oriented dynamics – formulated mathematically or not – is part of essentially every successful theory of the macroscopic world. And successful macroscopic theories should be taken as seriously as their microscopic counterparts. I shall now discuss how macroscopic entities are composed of microscopic building blocks and investigate whether goal-oriented dynamics can arise from goal-free microscopic laws.

As observed earlier, macroscopic entities displaying goal-oriented behavior are composed of a large number of microscopic building blocks, say, atoms or molecules. But not every collection of many atoms automatically becomes an agent which pursues goals. A piece of rock which we toss around is conveniently described in terms of the position and velocity of its center of mass (and in terms of its angular momentum) by mathematical equations similar to those of (classical) microscopic dynamics: deterministic, reversible, goal-free. This is due to the rock's rigidity. Since the relative positions of all constituting molecules stay the same when we toss it around, the dynamics of the constituents can be inferred from the dynamics of the whole body. The piece of rock is too rigid to become an agent. A macroscopic system can also be too flexible to become an agent. Think of a glass of water. The individual water molecules can essentially be displaced arbitrarily without changing the macroscopic state of the glass of water as a whole. And when we spill the water it ceases to be a glass of water but becomes a puddle. Goal-pursuing entities of a similar size, say, a frog or a mouse, are more flexible than the piece of rock but more rigid than the glass of water. Unlike the glass of water, they have to be rigid enough for some macroscopic features, like overall shape, to remain recognizable over a sufficiently long time. And unlike the piece of rock, they have to be flexible enough such that the dynamics of the constituents is not automatically determined by

[5] See, e.g., *Agent-based model*, `http://en.wikipedia.org/w/index.php?title=Agent-based_model&oldid=762339548` (last visited Feb. 17, 2017) as a starting point.

the dynamics of the whole, i.e. such that the microstate is not determined by the macrostate. Even without a more precise definition it is clear that there is much room for entities which are both, sufficiently rigid and sufficiently flexible.

Let me illustrate what I mean by saying that the macroscopic entities of interest have to be sufficiently rigid. To this end consider a mammal, say, yourself. Your body consists of head, neck, torso, and four limbs. Some macroscopic variables describing the state of your body would be position and velocity of your center of mass, and some angles and angular velocities for each joint. Now focus on your left arm. It consists of a very large number of biological cells, and each cell consists of a very large number of molecules. Not for all rearrangements of these molecules or even of the cells would the arm remain an arm. But the arm remains an arm, for a sufficiently long time, say, for some decades, and this is what I mean by saying that your body or your arm are sufficiently rigid. They consist of rigid subsystems. However, a given state of your left arm, say, 'forearm resting on the table', still corresponds to many different microstates, obtained by those internal changes after which the arm is still a functioning arm.

Here comes an important consequence of sufficient rigidity: There are as many different microstates corresponding to the macrostate 'forearm resting on the table' as there are microstates corresponding to other macrostates of your left arm, e.g., 'arm stretched out forward'. Therefore, the arm can evolve from one macrostate to another corresponding to a region of equal size in the space of microstates. This is different for the glass of water (or for the container of gas, which we discussed earlier) where different macrostates correspond to regions in the space of microstates of dramatically different size, cf. footnote 4. The macrostate of the glass of water essentially always evolves towards states corresponding to larger regions in the space of microstates. In contrast, the macrostate of a sufficiently rigid entity can evolve in several different ways. When I watch you with your left arm resting on the table, I describe the configuration of your body, your macrostate, in terms of the macroscopic variables introduced above. Based only on the knowledge of your macrostate, I cannot predict whether you will let your arm rest on the table for another few minutes or whether you will start stretching it out. At this point a mechanistic model of your macroscopic dynamics no longer works. Instead I will switch to a theory involving goals and intentions. When I see the glass of water for which your are reaching out, I no longer interpret the movement of your arm as a result of the macrostate in which it was before the movement started but as a result of your goal to pick up the glass.

You might object that my description of your body in terms of very few variables – position, velocity, some angles and angular velocities – was too crude. I could have described your macrostate in much more detail, e.g., including your breathing rate, the local appearance of your skin, visible tension of your muscles etc., and maybe also in terms of some not directly but in principle available variables like your heart rate or your blood pressure. Even with a much more detailed macroscopic description there is still a huge number of microstates corresponding to a given macrostate – you are still sufficiently flexible. And since you are also sufficiently rigid, different time evolutions of your macrostate are possible. This prevents me from formulating a good theory of your macroscopic dynamics simply in terms of the macroscopic variables. Instead I will resort to a theory including additional elements like goals and intentions. Theories of the latter kind are the most successful theories for describing and predicting, e.g., the behavior of animals or humans, their mutual interactions or the way in which we organize our society.

What have we achieved by this analysis from which we concluded that macroscopic entities featuring goal-oriented dynamics have to be sufficiently flexible (more flexible than the piece of rock) and sufficiently rigid (more rigid than the glass of water)? Have we derived goal-oriented dynamics from goal-free microscopic evolution, as Boltzmann has derived macroscopic irreversibility from reversible microscopic dynamics? Not quite. As noted above, our theories of goal-oriented dynamics are not equally concise as, say, thermodynamics or fluid mechanics. In order to put our understanding of goal-oriented behavior on a similar footing as our understanding of irreversibility, we would first have to formulate precise mathematical models for at least some simple examples of goal-oriented dynamics. Only then could we try to derive the equations for such models from microscopic theories invoking sufficient rigidity and flexibility. However, what our analysis shows is that, in contrast to what one might have expected, macroscopic theories containing elements of goal-oriented dynamics are not automatically at variance with goal-free microscopic laws!

Now we might once more ask, whether macroscopic entities of this type really display goal-oriented dynamics or whether this is only a feature of an approximation, whether the entities only seem to evolve in a goal-oriented way, whereas the real dynamics is goal-free mathematical evolution. As in the case of irreversibility I think that this question is ill-posed. Of course, if we describe a macroscopic entity in terms of all of its microscopic components, then this description will be goal-free. But then we no longer speak about said macroscopic entity at all, we only speak about the atoms or molecules which it is composed of. If we want to understand the behavior of the macroscopic entity as a whole, then we have to use a macroscopic language, i.e. seek a description in terms of macroscopic variables. For macroscopic entities which are sufficiently flexible and sufficiently rigid any good theory of this kind will contain goal-oriented dynamics. These macroscopic theories, including goals or intentions, are empirically valid, as we know from everyday observation. There is no reason to think of goal-oriented macroscopic behavior as less real than goal-free microscopic dynamics. And, as our analysis has shown, if the macroscopic entities at hand are sufficiently flexible and sufficiently rigid, then goal-oriented macroscopic dynamics need not be at variance with goal-free microscopic dynamics.

Ideas similar to the last argument have been developed also by other people in the context of free will,[6] where this position goes under the name of compatibilism. I make my point here in the general context of arbitrary goal-oriented behavior.

Summing up, we have seen that goal-oriented macroscopic dynamics is equally real as goal-free microscopic dynamics. Moreover, goal-oriented macroscopic behavior is compatible with goal-free microscopic laws, if the macroscopic entities under question are sufficiently flexible and sufficiently rigid. Under these circumstances mindless mathematical laws can give rise to aims and intention.

Acknowledgement: I thank Roderich Tumulka for helpful comments.

[6]See, e.g., Sean Carroll *Free will is as real as baseball*, http://blogs.discovermagazine.com/cosmicvariance/2011/07/13/free-will-is-as-real-as-baseball or George Musser *The universe is a big layer cake*, http://spookyactionbook.com/2015/09/03/the-universe-is-a-big-layer-cake/ for two popular accounts, and Christian List *Free will, determinism and the possibility to do otherwise* Noûs **48** (2014) 156–178, http://dx.doi.org/10.1111/nous.12019 for a more technical exposition.