# The concept of irreversibility: its use in the sustainable development and precautionary principle literatures

Dr. Neil A. Manson*

***Dr. Neil A. Manson** is an assistant professor of philosophy at the University of Mississippi.
Email: namanson =a= olemiss.edu (replace =a= with @)

Writers on sustainable development and the Precautionary Principle frequently invoke the concept of irreversibility. This paper gives a detailed analysis of that concept. Three senses of "irreversible" are distinguished: thermodynamic, medical, and economic. For each sense, an ontology (a realm of application for the term) and a normative status (whether the term is purely descriptive or partially evaluative) are identified. Then specific uses of "irreversible" in the literatures on sustainable development and the Precautionary Principle are analysed. The paper concludes with some advice on how "irreversible" should be used in the context of environmental decision-making.

**Key Words**

irreversibility; sustainability; sustainable development; the Precautionary Principle; cost-benefit analysis; thermodynamics; medicine; economics; restoration

## I. Introduction

The term "compassionate conservatism" entered public consciousness during the 2000 U.S. presidential campaign. It was created in response to the charge (made frequently during the Reagan era, and repeated after the Republican takeover of Congress in 1994) that Republicans lacked compassion – that they just did not care about the poor, the homeless, AIDS victims, and so on. George W. Bush felt Republicans had to peel off the "heartless" label if he was to win. The solution urged on him by political advisor Marvin Olasky was for Bush to market himself as a "compassionate conservative." Olasky had been promoting the concept for years, and he released *Compassionate Conservatism* early in 2000 (Olasky 2000). Bush often used the phrase on the campaign trail and after his election, and it appeared in the titles of numerous books and articles during and after the campaign. Yet even with Olasky's book, it was never very clear what a compassionate conservative was supposed to be – this despite Bush's assurance that he intended to govern as one. Much ink was spilled trying to figure out the meaning of this new-fangled political philosophy.

This effort was largely a waste of time. As so often happens in politics, a catchy term was introduced by "idea people" as a political slogan (notice the alliteration in "compassionate conservatism") with the content to be provided later. The whole point of a political slogan is to tap into some vaguely held sentiment. Political slogans are not meant to be sharply defined, for sharp definitions draw sharp boundaries and sharp boundaries force undecided voters out of one's constituency. To the extent that Bush and his supporters really did spell out compassionate conservatism as a coherent and detailed political program – a policy of federal support for religious charitable work and other "faith-based initiatives" – the term became less useful politically. The details alienated some conservatives (who saw it as a betrayal of their ideals regarding limited government) and angered most liberals (who saw it as violating the principle of the separation of church and state). That is why the Bush campaign, to the extent that it could get away with doing so, deliberately left the term "compassionate conservatism" vague. Those who sought the exact meaning of "compassionate conservatism" thus fundamentally misconceived the term's role. The most important thing to know about terms such as "compassionate conservatism" is not what they mean, but who uses them as if they were meaningful and what tensions their users are trying to hide.

I think the term "sustainable development" is in the same category as "compassionate conservatism" (I harbour similar misgivings about the term "the Precautionary Principle" – also alliterative).[1] "Sustainable development" was introduced and popularised as a political compromise term, designed to placate both the critics of environmentalism (who accuse environmentalists of being opposed to progress and development) and the critics of unbridled economic development (who view the "economic development first" mindset as the primary cause of global environmental destruction). As a political compromise term, vagueness and minimal content are of its essence. This is not necessarily a bad thing. Political compromise is sometimes noble, so if unification requires blurring some edges (synthesising antitheses, to use Hegelian language) then political compromise terms can serve a noble purpose. But the fuzziness of "sustainable development" does present a real problem for a philosopher trying to contribute to the first issue of *The Electronic Journal of Sustainable Development* – an issue devoted to addressing the nature of sustainable development. The obvious thing for a philosopher to do in a special issue about the nature of X is to explicate the concept of X. That cannot be done if X is an essentially vague concept. This is not to say that philosophers can have nothing to contribute to *subsequent* issues of this journal. They certainly can if those issues are devoted to more fruitful themes than the nonexistent "real nature" of sustainable development.

Instead, the time of environmental philosophers seeking to contribute to this journal is better spent taking a "ground up" approach. They should focus on concepts such as irreversibility, irreplaceability, irreparability, and catastrophe. It is through such concepts that notions such as sustainable development and the Precautionary Principle are typically characterised. Doing this work now will put future environmental philosophers (and others; philosophy is everyone's business!) in a much better position to provide precise technical definitions for "sustainable development" and "the Precautionary Principle" – definitions that make these terms more than just slogans. A close look at the use of the term "irreversible" in the literature on sustainable development and the Precautionary Principle will illustrate the approach. Before taking that look, however, a few more words are in order about the term "sustainable development".

## II. Misgivings about "Sustainable Development"

The term "sustainable development" gets attacked from at least two sides. Elements of both the environmental advocacy community and the free-market advocacy community agree that the environmentalist agenda is incompatible with a free-market economic system. Both also agree that the term "sustainable development" is a cover for something nefarious. What they disagree on is just what the term disguises. Is it, as some free-market advocates contend, merely a Trojan horse inserted into policy documents for the sake of the regulatory shackling of our vibrant global economy by technophobic and misanthropic environmentalists? Or is it a verbal sop tossed to Greens by international bodies, national governments, and multinational corporations as they continue to permit the destruction of the global ecosystem, as some environmentalists contend? Either way, both sides agree that the "official meaning" of "sustainable development" – the definition of it given in the Brundtland Report, say – is empty.[2]

Even many of those in the middle acknowledge the worry that "sustainable development" lacks meaning, or that it can mean almost anything to almost anyone. Consider this journal's self-description and its call for papers for this first issue (private communication). *The Electronic Journal of Sustainable Development* is described by its editors as "a new interdisciplinary journal that will publish peer-reviewed articles addressing policy, scientific, economic, technical and legal issues pertaining to 'sustainable development' (e.g. human well-being, economic growth, environmental quality, natural resource use and management, environmental regulation, technological change)." By that characterisation (particularly the "human well-being" clause), the editors will be hard-pressed to find an academic paper *not* about sustainable development. And in the call for papers, the editors say this of the Brundtland Report's canonical definition.

> "So defined, sustainable development is, like motherhood and apple pie, not a concept to which many would object. It would take a perverse outlook indeed to support the idea that people's needs should not be met both now and in the future. But, unobjectionable as it is in principle, the concept is sufficiently broad to allow a large spectrum of various interpretations."[3]

Analytic philosophers tend to think a concept that is both unobjectionable in principle and broad enough to tolerate radically different interpretations is worthless at best. In this case, we are probably not alone. The proverbial man-on-the-street, when presented with this description of sustainable development, will ask why anyone should passionately defend a concept to which no one would object.

Bryan Norton also worries that the term "sustainability" lacks clear meaning. Presumably that is why he entitled his recent collection of essays *Searching for Sustainability* (Norton 2003). He acknowledges (Norton 2003: 3) that the reason he chose the concept of sustainability as "the unifying concept to anchor normative theories of environmental protection" was that it "seemed non-controversial in many political contexts – perhaps because of its vagueness – but the concept seemed also to fit into a lot of conversations and even to act as a rallying point for diverse interests among those seeking a more environmentally sound way forward in environmental policy." In the same vein, he later says

> "It is no doubt useful, in policy discussions, to have a term like 'sustainability', which, like 'conservation' in days of old, can stand as a label for the many activities of environmentalists. The danger is that the term, like 'conservation' before it, will become a cliché. Nobody opposes it because nobody knows exactly what it entails." (Norton 2003: 168)

The parallel here between the function of the terms "sustainability" and "conservation" and the function of terms like "compassionate conservatism" is obvious. To be fair, Norton does call this situation a danger.

> "To avoid this trap it will be necessary for environmentalists, with the help of scientists and philosophers, to develop, explain and justify a theory of environmental practice that gives form and specificity to the goal of sustainability." (Norton 2003: 168)

Norton (perhaps unlike Bush) is not content to leave vague and undefined the concept he promotes. Nonetheless, the passages support my contention that even notable proponents of sustainable development acknowledge the term "sustainable development" is currently a catchall. Throughout his book Norton is explicitly engaged in the project of *providing* meaning to ("searching for") "sustainability". Why provide what is already present?

Norton tries to force form onto the protean term "sustainability". But that task presupposes a vocabulary with which the definition is to be framed. My complaint with the literatures on sustainable development and the Precautionary Principle is that many of the terms used to characterise these ideas themselves either are not clearly defined, are used equivocally, or are covertly given normative content. Some of these terms were mentioned earlier ("catastrophe", "irreparable", "irreplaceable"), but for the remainder of the paper let us focus on "irreversible". It pops up a lot in the literature on sustainable development and even more so in the literature on the Precautionary Principle, yet it appears in other, wholly unrelated literatures. Its use in these other fields may shed light on its meaning and use in discussions of sustainable development and the Precautionary Principle.

## III. Definitions of "Irreversible": Physical, Medical, and Economic

There are three specific academic disciplines in which the term "irreversible" is used as a technical term: physics, medicine, and economics. Indeed, it seems to me that all other technical senses of "irreversible" derive from the senses given to "irreversible" in those three fields; later I will give some examples of senses of "irreversible" that really reduce to one of these three senses. Let us see how "irreversible" is defined in each of these academic fields.

### *(1) Physics*

"Irreversible" has a clear theoretical definition in physics, specifically for use in thermodynamics. The thermodynamic sense of "irreversible" is intended for application in statistical mechanics, typically but not solely in connection with the definition of entropy (degree of disorder). This definition is usually given in the language of mathematical physics, but luckily for us philosopher of science K.G. Denbigh provides a non-mathematical, qualitative definition of "irreversible" that is sufficient for the purposes of this paper.

> "Reversibility and its negation are characteristics neither of 'things' nor of theories, but only of the *processes* which can occur in 'things'. Let us concentrate attention…on those macroscopic and inanimate 'things' which can be specified in terms of their temperature, volume and chemical composition, together with the intensities of any prevailing fields. Such specifications are sufficient to fix the momentary *macroscopic state* of the entity ('system') in question; a *process* is another temporal succession of such states due to the changing of one state into another.
>
> A process is said to be reversible if, and only if,

the system which undergoes that process, together with *all parts of its environment which are affected,* can be restored reproducibly to their original states. For example, let the system go from an initial state A, through states B, C, etc., to a final state X. The corresponding simultaneous states of the affected environment are α, β, χ etc. up to a final state ω. There is reversibility if it is possible not only for the system to be restored from X to A, but for this reversal to be accompanied by a simultaneous reversal of the affected parts of the environment from ω to α. In short all relevant parts of the universe must be capable of being put back how they were!" (Denbigh 1989: 506)

To visualise a process that reverses in this sense, think of running a movie backwards as in a Charlie Chaplin film, with broken bottles reassembling, anti-breezes putting sheets of strewn paper all back in a neat pile, and so on.

Obviously, we do not see physical processes reverse themselves very often, but certain physical systems under special conditions can be arenas for thermodynamic reversibility (or close approximations of it). For example, at low velocities, highly viscous fluids such as honey or glycerine demonstrate what is called "creeping flow" or "Stokes flow". Under these conditions, the motion of the fluid elements is reversible. If a moving boundary (e.g. a paddle) conveys an element of fluid (e.g. a drop of dyed glycerine) through a viscous medium (e.g. a vat of regular glycerine), the system exhibits Stokes flow when reversing the sequence and direction of the moving boundary reverses the trajectory of the fluid elements. In the sample experiment described, a twisting paddle apparently mixes the dye with the glycerine. Though the fluids appear mixed, they are really just stretched and folded to such an extent that from a distance the fluid seems to have an intermediate colouration of the dye. When the paddle is turned in the reverse direction to its original position, the dyed fluid elements "untangle" themselves, re-forming into a ball as if by magic.[4] If we blind ourselves to the fact that the experiment does not involve a closed system (a little bit of outside help is needed to twist the paddle), the Stokes flow experiment is a great illustration of thermodynamic reversibility.

Four things are worth noting about the definition of "irreversible" at work in physics. First, it specifies an ontology; that is, it specifies a realm of application for the term. On the physical definition, "irreversible" applies to *processes*, where the term "process" itself is defined in terms of "macroscopic state", and "macroscopic state" is defined in terms of sets of physical objects.

Second, the state descriptions are confined to the macroscopic level. The occurrence of an irreversible change in the system at, say, the molecular level will not disqualify a macro-scale process in the system from counting as reversible. Obviously, the more demanding one is regarding the state descriptions for a system, the harder it will be for some processes in that system to count as thermodynamically reversible. In the experiment discussed above, if the state of the fluid is specified down to the level of individual molecules, then even with Stokes flow the fluid has undergone irreversible change, since at that level there will be real, though perhaps indiscernible, shifts in the positions of the constituent molecules of glycerine even over short periods of time.

Third, "thermodynamically irreversible" carries no normative connotations. To say of a process that it is irreversible in the thermodynamic sense is not to say that there is something wrong with or bad about the process. This is fortunate, since, as Denbigh notes (Denbigh 1989: 507), "[c]omplete reversibility is not actually attainable in the real world. Irreversibility is the natural state of affairs, although the concept of reversibility remains a useful idealization for purposes of theory." Thermodynamic irreversibility is ubiquitous.

Fourth and finally, the features of some phenomena that lead us to describe those phenomena as "irreversible" in fact derive from thermodynamic irreversibility. We can regard all of those senses as special cases of the thermodynamic sense. For example, there is a sense in which any time-related phenomenon (e.g. any decision) is irreversible. The phrase "the arrow of time" refers to the fact that temporal events are ordered from past to future, with an essential asymmetry between past and future (the past is over and done with, whereas the future is open). The consensus view of philosophers seeking a physical rather than a metaphysical explanation of this asymmetry is that the future is the direction of time in which entropy (the degree of disorder) increases. What characterises entropy-increasing processes is that they are thermodynamically irreversible. If these philosophers are right, generic temporal irreversibility is really just a special case of thermodynamic irreversibility.[5]

In biology we find another example of a use of "irreversible" in which the phenomenon described is really just a special case of thermodynamic irreversibility: Dollo's Law.[6] The basic idea is that a structure or feature that has been lost or stripped away from an organism's ancestry during the process of evolution will not appear again in any of the organism's successors. But as Richard Dawkins points out, this is not a true law of nature – not a true exceptionless regularity.

> "'Dollo's Law' states that evolution is irreversible …[But it] is really just a statement about the statistical improbability of following exactly the same evolutionary trajectory twice (or, indeed, any *particular* trajectory) in either direction, but for larger numbers of mutational steps … the mathematical space of all possible trajectories is so vast that the chance of two trajectories ever arriving at the same point becomes vanishingly small … There is nothing mysterious or mystical about Dollo's Law, nor is it something we go out and 'test' in nature. It follows simply from the elementary laws of probability." (Dawkins 1986: 94)

Denbigh makes the point that his general physical definition of "irreversible" covers precisely such subsidiary laws, noting (Denbigh 1989: 504) that "in biology it is familiar enough that the evolutions of the various species of organisms do not normally occur in reverse. The feathered birds do not return to being scaly reptiles, nor do the reptiles revert to their own parent genera."

### *(2) Medicine*[7]

The medical/physiological definitions of "irreversible" and "irreversibility" are straightforward. In medical contexts, irreversibility is "the practical impossibility to heal disease", according to *Elsevier's Encyclopaedic Dictionary of Medicine, Part A: General Medicine*. "Irreversible" means "not capable of being reversed; characterizing a state or process from which recovery is impossible", according to *Blakiston's Gould Medical Dictionary*. Comparing this definition to the one for thermodynamic irreversibility, we notice several things. First, the ontology here is clearly different. In medicine, "irreversible" applies to *states of organisms* rather than to processes.

Some readers might be tempted to argue that ontologically there really is not a fundamental difference between medical irreversibility and thermodynamic irreversibility. As noted above, with thermodynamic irreversibility, processes are defined in terms of successions of macroscopic states. Should we say, then, that both the thermodynamic and medical senses of "irreversible" rest on an ontology of states? No. With thermodynamic irreversibility, macroscopic states are themselves specified in terms of the particular arrangement of particular macroscopic objects. If a change to a system removes or replaces any macroscopic components, the change automatically counts as irreversible – even if just one macroscopic component is replaced with an empirically indistinguishable duplicate. Hence the process of a stack of firewood burning is thermodynamically irreversible, even if we keep the fire going indefinitely by replacing burnt logs with qualitatively identical fresh logs. With medical irreversibility, on the other hand, states of organisms are *not* specified in terms of the particular arrangements of particular organic objects, but rather are defined functionally. For example, a person who has had her skin cut deeply is not in an irreversible condition. Even though a scar will result from the healing, her skin can reacquire all of its old functions, so the damage done by the cut is reversible. Yet the healing process – the process of clotting blood, providing new skin, and so on – does not count as the thermodynamic reversal of the act of skin-cutting. Let us mark this important distinction by saying that thermodynamic irreversibility is a *strict identity* concept whereas medical irreversibility is a *functional* concept.

Second, there is a normative element to the medical definition of "irreversible", insofar as health, disease, and recovery are normative concepts. To say of an organism that it is healthy is to say that it is the way it should be, that it is in the state that is normal or appropriate for organisms of its kind. Disease is a normative concept because health is a normative concept, and disease is defined in terms of health. The same goes for the concept of recovery.

Lastly, note that medical irreversibility is not just loss of homeostasis. Homeostasis is the state wherein an organism maintains itself in response to environmental changes. If an organism is in a state of irreversible decline or disease, then that organism is no longer in homeostasis. The converse is not true, however, because homeostasis is not indexed to the current possibilities for technological intervention, whereas medical reversibility is. For example, a person with a gunshot wound to the heart is no longer in homeostasis, but may be in a medically reversible state thanks to advances in trauma technologies.

In ecology, the concept of homeostasis is applied not just to individual organisms, but to larger systems, from small, isolated ecosystems, to species, all the way up to the Earth as a whole. Might we extend the concept of medical irreversibility to larger systems as well? If we conceive of, say, the Earth as a whole as an organism ("Gaia"), might we sensibly speak of its prospects for healing or recovery? Of irreversible damage to it? Nothing prevents this extension, but we should be aware that it rests on an assumption not all parties to disputes about the environment accept: that ecosystems, the Earth, and so on are properly conceived as organisms.

If the extension is only metaphorical, then it is unclear whether the normative connotations of the strict medical use carry over to the metaphorical use. Even supposing the extension is literal, we should still be careful to distinguish this extended sense of medical irreversibility from the loss of homeostasis. It *may* be the case that, for large ecosystems or the Earth as a whole, loss of homeostasis is necessarily equivalent to suffering irreversible damage, simply because no human technological interventions to restore homeostasis are possible once systems so large fall out of homeostasis. But that equivalence needs to be *argued for*, not just assumed. Certainly it is conceivable that technological interventions on a grand scale can restore homeostasis to very large systems, in which case the system would have suffered homeostasis-disrupting damage but not irreversible damage (in the extended medical sense of "irreversible" just proposed). Just because there is a threat of homeostasis disruption to a system, it does not follow that there is a threat of irreversible damage to that system.

### *(3) Economics*

The use of the terms "reversible" and "irreversible" in economics seems largely restricted to environmental economics.[8] Even in this field, it seems that not much work has been done to clarify the concept of economic irreversibility.[9] The attempts I have seen all involve Anthony Fisher. In collaboration with various authors over the years, he has tried to articulate a sense of "irreversible" that would apply in economics, specifically in connection with issues of the use of natural resources.[10] An extended account of irreversibility in environmental economics can be found in a piece he co-authored with John Krutilla (Fisher and Krutilla 1985). What follows is a synopsis of their account.

They begin by observing that natural landscapes and ecosystems took so long to form that humans cannot replace or restore them "in all their essential features" after they are destroyed. "There is thus a basic irreversibility that attends the modification of unique scenic or biological environments" (Fisher and Krutilla 1985: 173). They deny that this basic irreversibility is just the irreversibility attaching to any time-related decision (the kind of irreversibility that many philosophers of physics think derives from thermodynamic irreversibility). Rather, the kind of irreversibility at issue concerns the degree to which the effects of a poor decision can be remedied. For a completely irreversible decision, the effects cannot be remedied to any degree whatsoever. For example, a decision to use a plant to produce a certain inventory is much more reversible than a decision regarding the very production capacity of a plant. Excess inventory can be liquidated and future production can be scaled back or redirected, but the lost investment in excess capacity cannot be recouped. The same dynamic applies to decisions to develop natural resources. For example, the decision to use the geysers at Yellowstone National Park for geothermal energy would have a very low degree of reversibility, just because the original geysers would be nearly impossible to restore.

Why is it so important to avoid taking decisions that will, if they go wrong, require environmental restoration? The authors give two reasons. First, restoration is typically difficult if not impossible, and the losses of value that cry out for restoration typically last long periods of time – at least, long relative to "the time span that is meaningful to human societies" (Fisher and Krutilla 1985: 176). Second, even with restoration, there is a permanent loss of "authenticity." Just as no forgery, however perfect, is an adequate substitute for an original artwork, so is no environmental restoration an adequate substitute for the original – at least among "purist" wilderness seekers. Whatever the objective merits of these demands for authenticity and purity, they are felt with great intensity by a significant submarket of outdoor recreation enthusiasts. In the calculations of economists, this translates into disfavouring irreversible decisions to develop natural resources.

As with the thermodynamic and medical senses, the economic sense of "irreversible" has a distinct ontology. Economic irreversibility is a feature of *decisions*. And irreversibility in this sense is partly a normative concept. Other things being equal, a decision which is reversible is better than – is to be favoured over – a decision which is irreversible. Other things being equal, the more reversible a decision is, the better. Notice also that, unlike the thermodynamic and medical senses of "irreversible", in the economic sense reversibility and irreversibility come in degrees.[11]

They admit (Fisher and Krutilla 1985: 175) that "[a]ny investment in a specialized plant and equipment will, in a sense, represent an irreversible commitment of capital to an undertaking." And since the case of investment in plant and equipment is just an illustration of a more basic feature of economics, they admit there is a sense of "irreversible" in which irreversibility is ubiquitous in economics. This will not be the trivial thermodynamic, you-can-never-go-back-in-time sense, since they have already denied that that is the kind of irreversibility they have in mind. Presumably they are thinking of the ubiquity of the lack of a guarantee that losses from capital investment will be recouped. As is made clear below, if a feature is ubiquitous, then it cannot figure into any of our practical deliberations, and hence cannot figure into economics. This suggests something entirely different is central to their definition of "irreversibility" – namely, the impossibility of restoration. On their definition, to say that a decision to develop some natural resource is completely irreversible is to say that upon completion of the development the environment affected by that development will be incapable of being restored, and to say that that decision is reversible to some degree is to say that the affected environment can be restored to some degree.

If this interpretation of Fisher and Krutilla is correct, then we immediately face a dilemma. If they are conceiving of natural landscapes and ecosystems as organisms, and if by "restore' they just mean "return to health" or "return to homeostasis", then really they are not introducing a new sense of "irreversible". They are just taking the (extended) concept of medical irreversibility and porting it over to economics. In that case, they would not be arguing for anything new or profound when they argue that there are strong economic reasons to avoid irreversible decisions, because they would just be arguing that there are strong economic reasons to avoid decisions that lead to death. Economists (and the rest of us!) already knew that. The only contributions environmentalist thinking lends to this issue are the arguments that various non-human organisms (sentient animals, all animals, all living things) are things whose death counts and that various seeming non-organisms (species, ecosystems, Gaia) are sufficiently similar to organisms in the relevant respects to deserve consideration equal to that afforded organisms. On this construal, then, there is not a distinctly *economic* sense of "irreversible" – just an importation into economics of the extended medical sense of "irreversible".

On the other hand, if by "restore" they mean something broader than or different from the extended medical sense of "irreversible" – say, whatever is meant in the (vast) literature on environmental restoration – then there seems to be little point in adding "irreversibility" to the environmental economist's vocabulary. Any reference to irreversible decisions could be replaced with reference to decisions such that the environmental systems or objects affected by them cannot be restored. If that is what is meant by "irreversible decision", it would be better just to spell that out each time "irreversible decision" is used. That way we could avoid the temptation to conflate irreversibility-as-unrestorability with medical or thermodynamic irreversibility. (More generally, for the sake of clarity and rigorous discussion in any field, it is useful to reduce the number of synonyms in the disciplinary vocabulary; in the ideal case, each term in the theoretical language will have one and only one sense, and will serve one and only one function.)

Whether the foregoing interpretation of Fisher and Krutilla is correct and, if so, which horn of the dilemma it is better to grab are issues left for the reader. What we can say for sure is that there is a sense of "irreversible" which some environmental economists think is distinct to the field of economics, that this sense applies to decisions, and that in this sense economic irreversibility is a normative concept.

## IV. The Inconsistent Use of "Irreversible" in the Environmental Literature

So far we have seen there are three distinct senses of "irreversible". Each sense can be characterised with respect not just to its academic field, but to its ontology (the proper realm of application for that sense of "irreversible") and with respect to what we might call its normative status (whether a predication of that kind of irreversibility is purely descriptive or partly normative). The chart below spells out the options.

Unfortunately, many of those who write about or think about environmental problems either use "irreversible" in none of the above three senses, or use it equivocally, or use it in some vague and confused sense that is a mish-mash of the above three senses. This is unacceptable if the notion of irreversibility is going to figure into any serious academic or policy discussion.

Consider Principle 15 the 1992 Rio Declaration of the

**Three Kinds of Irreversibility**

| *name* | *ontology* | *normative status* |
|---|---|---|
| thermodynamic irreversibility | physical processes | purely descriptive |
| medical irreversibility | states of organisms | partly normative |
| economic irreversibility | decisions | partly normative |

U.N. Conference on Environment and Development, which provides a canonical definition of the Precautionary Principle.

> "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."

No definition of "irreversible" is provided in the Rio Declaration. In context, it is not clear what sense of "irreversible" is intended: the thermodynamic sense, the medical sense, the economic sense, or some other sense. This just will not do. If we do not get from the Rio Declaration (or supplements to it) some guidelines for what it is for damage to be irreversible, we cannot know whether to apply the Precautionary Principle.

The lack of clarity and consistency in the use of "irreversible" makes it hard for critics of concepts such as sustainable development or the Precautionary Principle to hit their targets. Indur Goklany (Goklany 2001) attempts to reformulate the Precautionary Principle in such a way that it does not promote counterproductive regulation of technology. Setting out criteria for application of this reformulated Precautionary Principle, he suggests the following.

> "[When] the action under consideration results in positive as well as negative environmental impacts unrelated to public health," one of five criteria to use is "the irreversibility criterion. Greater priority is to be given to outcomes that are irreversible, or likely to be more persistent." (Goklany 2001: 9–10)

Here Goklany is just trying to capture part of what he thinks its proponents mean by "the Precautionary Principle". Yet "irreversible" is nowhere defined in the text, putting Goklany in the position of criticising a criterion without knowing just what the criterion is.

We get a clue as to what he thinks proponents of the Precautionary Principle have in mind when they use "irreversible" from the handful of times he appeals to the irreversibility criterion. Regarding the argument that we should phase out the use of the pesticide DDT because "DDT accumulation in the environment may lead to irreversible environmental harm", he responds (Goklany 2001: 21) that "the death of a human being is as irreversible as, and more heinous than, the death, for instance, of a bird. That is, there is no moral equivalency between the two outcomes." Here Goklany seems to think proponents of the Precautionary Principle have narrow (not extended) medical irreversibility in mind when they advance the irreversibility criterion, but that they count nonhuman animals among the things that can be harmed. That is, he thinks "irreversible" as used by proponents of the Precautionary Principle just means something like "being such as to cause the death of particular organisms".

Yet surely such proponents have a stronger objection in mind than just that, once released, DDT will take many animal lives, for then applying the irreversibility criterion would just be a matter of weighing the human lives saved by a technology against the nonhuman lives lost. Standard cost-benefit analysis is perfectly capable of doing that kind of weighing. The real fight would simply be whether to count nonhuman animals in our cost-benefit calculations. But the whole point of the Precautionary Principle (and its particular clauses such as the irreversibility criterion) is to *delimit* cost-benefit analysis – to say that, when certain criteria are met, we *cannot* simply weigh the costs against the benefits. The proponents of the Precautionary Principle are not just saying that standard cost-benefit analysis does not assign value to nonhuman animals. They are saying that, *even if it does*, cost-benefit analysis is still a defective procedure because it neglects important features of decisions – for example, that the consequences of the decision are irreversible. But just what is that other sense of "irreversible" – the sense whereby the irreversibility criterion is not an injunction to avoid killing organisms? Goklany has certainly read a great deal of the literature on the Precautionary Principle. Perhaps there is a clear, consistent use of the term "irreversible" in this literature that he just fails to discern. More likely, however, is that the misinterpretation by Goklany is explained by the confused use of "irreversible" in the literature. In any case, it is not Goklany's job to make sense of the irreversibility criterion. In this context, it is the job of the proponents of the Precautionary Principle.

For another example, consider this criticism of the Precautionary Principle from Julian Morris.

> "…all change (and hence all damage) is irreversible in the strict sense that the precise structure of the world that pertained before cannot once again come into being. This is a consequence of the second law of thermodynamics, wherein it is observed that the state of disorder (or entropy) of

> the universe is constantly increasing. [...] This ultimately negates the utility of including 'irreversible' as a criterion distinct from 'serious'." (Morris 2000: 14)

Here Morris attributes use of the thermodynamic sense of irreversibility to proponents of the Precautionary Principle. Surely that is not what they have in mind, given that thermodynamic irreversibility is ubiquitous and is neither good nor bad in itself. So why the misattribution by Morris? It derives, I suggest, not from uncharitable interpretation on his part, but from the paucity of explicit definitions of "irreversible" in the literature on the Precautionary Principle. From this philosopher's perspective, Morris is being a charitable interpreter because he is attributing a well-defined use of "irreversible" to proponents of the Precautionary Principle. It is more charitable to attribute a mistaken but well-defined use of a term to a person than it is to attribute an unmistaken-because-undefined use of a term to that person. At least in the former case, the person being criticised says something false. In the latter case, the person being criticised does not even say anything.

Lastly, consider how the respected environmental philosopher Holmes Rolston III uses the term "irreversible" in stating what he calls "The Reversibility Maxim: Avoid Irreversible Change".

> "We do business in a many-splendored natural system, one where life has so far prospered. It vastly exceeds our mastery, is incompletely understood still, and its mysterious origins and dynamics are perhaps finally unfathomable. All evolution is irreversible but moves very slowly. Here humans want to avoid precipitous irreversible changes, or even minor ones we later regret." (Rolston III 1989: 155)

The first use of "irreversible" is most reasonably understood as referring to Dollo's Law.[12] As I argued earlier, this sort of evolutionary irreversibility is really just a special case of thermodynamic irreversibility. Interestingly, Rolston has elsewhere argued that the process of evolution itself is intrinsically valuable.[13] Since that process is thermodynamically irreversible, we have at least one case of an environmental philosopher holding that a certain irreversible process is intrinsically good! Since thermodynamic irreversibility is ubiquitous, however, it is unavoidable, and so Rolston cannot (or, at least, should not) be using "irreversible" in that sense in his second use of the term. Here it is not clear exactly what sense of "irreversible" he has in mind. Since decisions are proper objects of regret, perhaps he is using "irreversible" in the economic sense in the second instance. Rolston's use here seems to be either confused or equivocal – a fact we come to realise only after having clarified exactly what senses "irreversible" possesses.

## V. How to Use "Irreversible": Some Suggestions

At this point we can draw some lessons about the meaning and use of "irreversible" in the context of environmental decision-making. First, we can identify three mistakes in reasoning invited by the existence of the three different senses of "irreversible".

The most obvious mistake in reasoning is *equivocation*. Equivocation is the use of the same term in different senses in order to carry an argument to its conclusion – for example, "Something is better than nothing, and nothing is better than the sweet potato cheesecake at Galatoire's, so something is better than the sweet potato cheesecake at Galatoire's." Equivocating can lead us from true premises to a false conclusion (for example, that something is better than the sweet potato cheesecake at Galatoire's), so it is not a valid form of argument. Given the multitude of senses of "irreversible", readers of the literature on environmental decision-making should be on the lookout for equivocal uses of the term.

The second mistake in reasoning is *deriving a normative conclusion from purely descriptive premises*, also known as "deriving an 'ought' from an 'is'." If the conclusion of an argument is that some thing is right or ought to be done, or is wrong or ought not to be done, then amongst the premises there must be one about what is right or ought to be done, or what is wrong or ought not to be done. David Hume famously identified the error (Hume 1739 [1978]), G.E. Moore confirmed it and labelled it "the naturalistic fallacy" (Moore 1902 [1988]), and almost all philosophers – with exceptions such as John Searle (Searle 1964) – have agreed that it is, indeed, a fallacious form of reasoning.[14] In the context of our discussion, the point is fairly basic. If we are going to reason from the irreversibility of something to the conclusion that we ought to avoid that something, then somewhere in the premises of the argument it must be asserted that (a) the irreversibility of that something is bad, disvaluable, or to be avoided, or (b) the irreversibility of that something, in conjunction with some other normative premise, implies that some other thing is bad, disvaluable, or to be avoided. Otherwise, the fact that something is irreversible cannot tell us what we *ought* to do at all.

One thing we see directly upon recognising this mistake is that no normative conclusion can be derived from the fact that a certain process is thermodynamically irreversible. Thermodynamic irreversibility is not a normative concept, but a purely descriptive one. Since, in the context of environmental decision-making, the notion of irreversibility is almost always invoked in the course of telling us what we ought to do, it follows that *thermodynamic irreversibility is irrelevant to environmental decision-making*. If "irreversible" is used in such contexts, the usage must convey some sense other than the thermodynamic one. Note that this also applies to any senses of "irreversible" that derive directly from thermodynamic irreversibility – for example, the irreversibility of evolutionary processes noted in Dollo's Law.

The third mistake in reasoning derives from another philosophical injunction regarding "ought": "'ought' implies 'can'." If we think a person *ought* to do something, we must think the person *can* do that thing. So if a person cannot do a certain thing, then we cannot hold that the person ought to do that thing. If I see a city bus crushing a child under its front wheel, there are various things I ought to do: alert the bus driver, call the police, run for a doctor. But it is not the case that I ought to flip the bus over to relieve the pressure on the child. Since I cannot do that thing, it cannot be that I ought to do it. For a more topical example, consider Rolston's Irreversibility Maxim: "Avoid irreversible change." If all changes are irreversible (in whatever sense of "irreversible" is at work), and if we cannot avoid changing things, then we cannot avoid making irreversible changes, and therefore it cannot be that we ought to avoid making irreversible changes. The lesson here is that, if irreversibility is ubiquitous in the realm of application of a particular sense of "irreversible", then the fact that something is irreversible cannot tell us what we ought to do at all, and so the fact that something is irreversible cannot figure into our practical deliberations.[15] Let us dub this third mistake *the fallacy of ubiquitous irreversibility*. Obviously, whatever sense we give to "irreversible" when we are talking about the environment, it should not be one whereby every harm, or every decision we take, or every process whatsoever (or whatever else is the intended realm of application of "irreversible") is irreversible.

This obviously rules out defining "environmental irreversibility" as Denbigh defines "thermodynamic irreversibility". I would like to end this paper by considering a few specific ways in which the environmental sense of "irreversible" – whatever sense we decide to give it – should differ from the thermodynamic sense of "irreversible". First, the environmental sense should be one in which "borrowing" from the environment is allowed. A simultaneous reversal of the parts of the environment affected by a process is required for the process to be thermodynamically reversible. This demand will have to be dropped for environmental reversibility. If not, then building a hydroelectric dam (for example) will automatically count as doing irreversible damage. After all, removing the dam will require, among other things, gasoline, explosives, bulldozers, cranes – in short, materials and energy that are not parts of the dam. Drawing on those materials and that energy will require changing some part of the environment. Since any attempt to return some part of the world to its original state will involve drawing on resources from some other part of the world, anyone who insists that such damage counts as irreversible will be committing the fallacy of ubiquitous irreversibility.

Second, there should be no demand that, for a change to the environment to count as environmentally reversible, the pathway of restoration must retrace in reverse the pathway of destruction. Otherwise we will be committing the fallacy of ubiquitous irreversibility once again. For example, the process of putting up a dam will involve shutting off the flow of water upstream, then building a dam slab by slab from the bottom up, then restarting the flow of water upstream. The process of bringing down the dam, however, will require, not taking apart the dam slab by slab, but blowing it up with explosives. Only rarely do the ways of returning nature to its unmodified state involve a complete running-in-reverse of the original process. (About the only technique that comes to mind is standing back up in order to repair the damage done by having sat down on a particularly springy bush.)

Lastly, great care should be exercised in developing criteria for permissible state descriptions. Remember that thermodynamic irreversibility is a feature of processes, where processes are simply successions of states of a system, and system states, in turn, are simply sets of macroscopic, observable objects with specific properties. In the Stokes flow experiment described earlier, the process of twisting the paddle amidst the fluid counts as reversible only because the state description is macroscopic. There is no demand that each individual molecule of glycerine ends up in the same place in the tub. If there were such a demand, not even the process exhibited in the Stokes flow experiment would count as reversible.

Instituting such a demand as part of a definition of environmental reversibility would, once again, involve commission of the fallacy of ubiquitous irreversibility. We can never put things back the way they were, molecule-for-molecule. If not molecule-for-molecule, however, what level of detail is appropriate? For example, when evaluating whether the damage caused by a dam is irreversible or not, an important factor will be the level of detail of the description of the affected area prior to the construction of the dam. The state description could be fairly minimal – say, "fresh water flows atop a sloping bed of assorted rocks and boulders, with fish swimming about." But if the state description is more specific – say, "fresh water flows at a rate of one meter per second atop a bed of granite rocks and boulders at a slope of one degree, with rainbow trout and pacific salmon swimming about" – then obviously it will be harder to return to that state, and hence easier for the damage done to count as irreversible. Without criteria regarding permissible state descriptions, unscrupulous sorts might classify any damage whatsoever as irreversible, just by describing the original state of the environment to an arbitrary degree of precision.

Particular care must be exercised when it comes to allowing causal or genetic properties into one's state descriptions. For example, one would think that in disassembling a valuable artefact – say, a Studebaker Golden Hawk – the owner will not have done any irreversible damage to it, since she can reassemble the car later on. Yet the beginning car will have the property of having been assembled by the workers of the Studebaker Corporation, while the reassembled car will not. Instead, it will have the property of having been assembled by the owner. If such properties as having been assembled by Studebaker employees are permitted to be part of the description of the original state of the Studebaker, then the mere act of disassembly irreversibly changes the car. Likewise, if properties such as being pristine are allowed into the descriptions of the environment, then it will become trivially true of any given environment that any damage to it is irreversible, and once again we will be committing the fallacy of ubiquitous irreversibility. To use the terminology introduced earlier for distinguishing the ontology of medical irreversibility from the ontology of thermodynamic irreversibility, strict-identity state descriptions will make environmental irreversibility ubiquitous, while merely functional state descriptions will not.

This is not to deny any place to causal or genetic properties in our thinking about the environment. Species are defined by their histories as much as by the physical characteristics (colouration, wing shape, etc.) of their members.[16] If a planned development threatens destruction of a species, that is typically considered a strong, if not decisive, reason for stopping it. But in that case, the threat to species is reason enough to stop the proposed development. The prospect of species loss should not be counted against the development twice – once as the loss of species, then a second time under the guise of threatening irreversible damage. Likewise, some philosophers (Elliott 2003; Katz 2003) criticise the very idea of environmental restoration as being incoherent. They argue that any restored environment is actually an artefact, not a natural object. If they are right, then even creating a perfect replica of a given pre-development environment will not count as having restored that environment, just because the replica has the causal/genetic property of having been created by humans. Whatever its merits, this view of restoration should not be covertly presupposed in anyone's state descriptions. The "true restoration is impossible" thesis should be kept separate from one's account of environmental irreversibility and stated explicitly. Covertly folding the view that true restoration is impossible into one's notion of environmental irreversibility merely invites confusion and semantic disputation.

## VI. Conclusion

Environmental irreversibility may not be an indispensable concept in environmental philosophy and environmental decision-making, but as I hope the preceding discussion has shown, it plays enough of a role to merit scrutiny. To deserve the place it has, the concept needs to be defined rigorously and distinguished clearly from other concepts at work in our thinking about the environment. If we cannot do so, we should avoid talking about environmental irreversibility altogether.

## Notes

1. Per Sandin (Sandin 1999) and I (Manson 2002) see a thicket of competing formulations of the precautionary principle. In my view, the formulations range from the vacuous to the self-refuting. It is possible there exists a coherent, precise, palatable, and useful version of the Precautionary Principle that is also inconsistent with the principles of standard cost-benefit analysis (for all of its vagaries, that is the one thing the Precautionary Principle is supposed to be – an alternative to standard cost-benefit analysis). There is some debate over whether there really is such a version. In addition to the papers cited above, see Turner and Hartzell (2004) and Sandin (2004).
2. Officially (World Commission on Environment and Development 1987: 43), "[s]ustainable development is development that meets the needs of the present without

compromising the ability of future generations to meet their own needs."

3. The "motherhood and apple pie" and "perverse outlook indeed" comments suggest the editors are in sly agreement with the aforementioned suspicious environmentalists and free marketers that "sustainable development" is a sugar-coated phrase whose advocates are up to no good.
4. Many thanks to Dr. Ken Kiger of the University of Maryland's Department of Mechanical Engineering for providing me with the definition of Stokes flow and the details of the experiment illustrating it. Kiger also directed me to a demonstration of this experiment narrated by G.I. Taylor and available online at http://web.mit.edu/fluids/www/Shapiro/ncfmf.html. Click on the video titled "Low Reynolds Number Flow". The experiment is shown starting around minute 13.
5. For a detailed presentation of the argument that time's arrow is explicable in terms of thermodynamic irreversibility, see Chapter 3 of Sklar (1992).
6. I would like to thank Derek Turner for calling Dollo's Law to my attention.
7. I would like to thank the audience at my reading of a much earlier version of this paper for calling to my attention the medical use of "irreversible". The earlier version was "The Concept of Irreversibility in the Application of the Precautionary Principle", read at the Society for Philosophy and Technology's July 2003 meeting in Park City, Utah.
8. An unscientific survey of general economic dictionaries revealed only one in which "irreversibility" was defined. Even there, "irreversibility" was only defined by example. See Black (2002).
9. The term "irreversible" does not appear in the paper title or list of key words for any article in *The International Journal of Sustainable Development* since its inception in 1998, nor in any paper title in the journal *Sustainable Development* during that same time. (*Sustainable Development* does not require its contributors to provide key words in connection with the abstracts.) These are (only currently, we hope) the two leading journals regarding sustainable development; many of the pieces they publish are in the area of environmental economics. If irreversibility were a foundational idea in environmental economics, one might expect formal reference to it in these journals. Hence we can take the absence of such reference as (weak) evidence that irreversibility is not a foundational idea.
10. The earliest of his efforts was with Kenneth Arrow. See Arrow and Fisher (1974).
11. Bryan Norton also treats irreversibility as a degree (or "scalar") concept. See figure 12.1 (Norton 2003: 211).
12. I thank Derek Turner for suggesting Rolston had Dollo's Law in mind.
13. I heard Rolston say so at his talk "Naturalizing and Systematizing Evil," delivered at the Gifford Bequest International Conference – "Natural Theology: Problems and Prospects" – held at the University of Aberdeen, Scotland in May 2000.
14. For a response to Searle, see Thomson and Thomson (1964).
15. For the record, I do not deny that, as a purely philosophical matter, it is coherent to maintain that a certain feature of reality is both ubiquitous and ethical/evaluative in character. For example, I do not deny that it is coherent to maintain that all natural objects are good – that Nature as a whole, from galactic clusters on down, is thoroughly suffused with goodness. However, if it really is the case that goodness is ubiquitous, then this fact cannot figure into our practical deliberations with respect to natural objects. It drops out of the equation.
16. Indeed, some philosophers of biology claim that if a perfect replica of some earthly organism existed on another planet, that replica still would not be a member of the earthly species. Elliott Sober (Sober 2000: 151) says that "if we discovered that other planets possess life forms that arose independently of life on earth, those alien organisms would be placed into new species, regardless of how closely they resembled terrestrial forms. Martian tigers would not be tigers, even if they were striped and carnivorous."

## References


Arrow, K.J. and A.C. Fisher. 1974. "Environmental Preservation, Uncertainty, and Irreversibility." *Quarterly Journal of Economics* 88 (2): 312–319.

Black, J. 2002. *A Dictionary of Economics*, 2nd edition. New York: Oxford University Press.

Dawkins, R. 1986. *The Blind Watchmaker: Why the evidence of evolution reveals a universe without design*. New York: W.W. Norton & Company.

Denbigh, K.G. 1989. "The Many Faces of Irreversibility." *The British Journal for the Philosophy of Science* 40 (4): 501–518.

Elliott, R. 2003. "Faking Nature." In *Environmental Ethics: An Anthology*, edited by A. Light and H. Rolston III. Malden, Massachusetts: Blackwell Publishers.

Fisher, A.C. and J.V. Krutilla. 1985. "Economics of Nature Preservation." In *Handbook of Natural Resource and Energy Economics: Volume I*, edited by A.V. Kneese and J.L. Sweeney. New York: Elsevier Science Publishers B.V.

Goklany, I. 2001. *The Precautionary Principle: A Critical Appraisal of Environmental Risk Assessment*. Washington, D.C.: Cato Institute Press.

Hume, D. 1739 [1978]. *A Treatise of Human Nature*, edited by L.A. Selby-Bigge. New York: Oxford University Press.

Katz, E. 2003. "The Big Lie: Human Restoration of Nature." In *Environmental Ethics: An Anthology*, edited by A. Light and H. Rolston III. Malden, Massachusetts: Blackwell Publishers.

Manson, N. 2002. "Formulating the Precautionary Principle." *Environmental Ethics* 4 (3): 263–274.

Moore, G.E. 1902 [1988]. *Principia Ethica*. Buffalo, New York: Prometheus Books.

Morris, J. 2000. "Defining the Precautionary Principle." In *Rethinking Risk and the Precautionary Principle*, edited by J. Morris. Woburn, Massachusetts: Butterworth-Heinemann.

Norton, B. 2003. *Searching for Sustainability: Interdisciplinary Essays in the Philosophy of Conservation Biology*. New York: Cambridge University Press.

Olasky, M. 2000. *Compassionate Conservatism: What It Is, What It Does, and How It Can Transform America*. New York: Free Press.

Rolston III, H. 1989. *Philosophy Gone Wild*. Buffalo, New York: Prometheus Books.

Sandin, P. 1999. "Dimensions of the Precautionary Principle." *Human and Ecological Risk Assessment* 5 (5): 889–907.

Sandin, P. 2004. "The precautionary principle and the concept of precaution." *Environmental Values* 13 (4): 461–475.

Searle, J. 1964. "How to Derive 'Ought' from 'Is'." *Philosophical Review* 73 (1): 43–58.

Sklar, L. 1992. *Philosophy of Physics*. Boulder, Colorado: Westview Press.

Sober, E. 2000. *Philosophy of Biology*, 2nd edition. Boulder, Colorado: Westview Press.

Thomson, J. and J.J. Thomson. 1964. "How Not to Derive 'Ought' from 'Is'." *Philosophical Review* 73 (4): 512–6.

Turner, D. and L. Hartzell. 2004. "The Lack of Clarity in the Precautionary Principle." *Environmental Values* 13 (4): 449–460.

World Commission on Environment and Development. 1987. *Our Common Future*. New York: Oxford University Press.