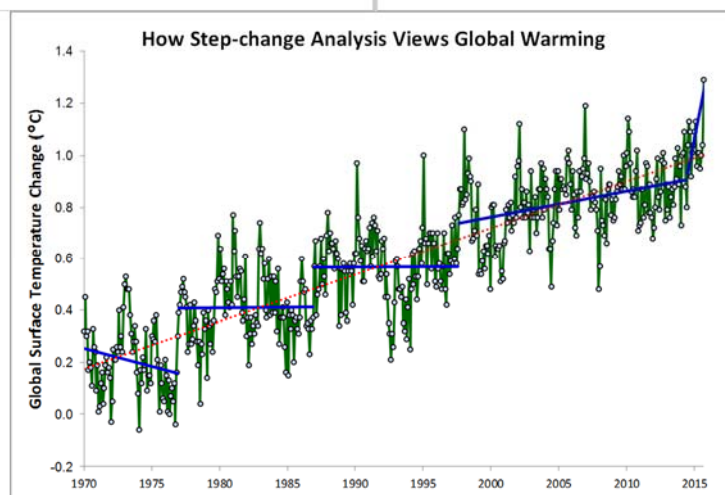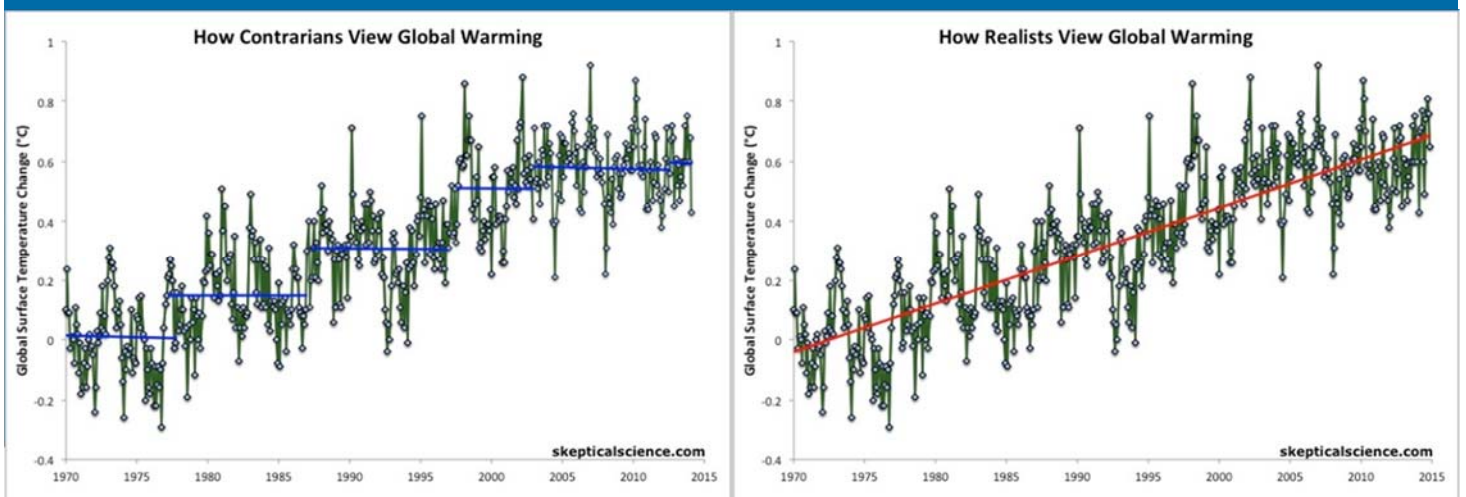# The climate wars and "the pause" – are both sides wrong?

Roger N Jones and James H Ricketts

Climate Change Working Paper 37

Victoria Institute of Strategic Economic Studies

Key words: air temperature, global warming, decadal variability, climate change, philosophy of science, scientific theory

Suggested citation

Jones, R.N. and Ricketts, J.H. (2016) The climate wars and "the pause" – are both sides wrong? Climate Change Working Paper No. 37, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, Australia.

Publishing Notes

This is a pre-submission working paper and a companion paper to Reconciling the signal and noise of atmospheric warming on decadal timescales, Climate Change Working Paper. No. 38.

Abstract

This paper takes as its starting point the debate surrounding whether global warming after 1998 is a routine part of a gradual, long-term trend or whether discontinuities in the temperature record either show the theory is wrong or that climate risks are overstated. It argues the main scientific defence, though correct, is scientifically weak and open to criticism because it sidesteps the role of nonlinearity in the warming process. Additionally, the distinction between short- and long-term trends is subjective and depends on the method used to create that distinction. The paper applies three areas of the philosophy of science to reconcile the situation: (1) it uses the structure of scientific research programmes to identify the key core and auxiliary global warming theory in order to frame a more robust scientific response; (2) it addresses theory falsification and confirmation to identify spurious arguments aimed at discounting global warming theory based on dogmatic empiricism and (3) it investigates the potential for severe testing to investigate scientific hypotheses supporting gradual trend-like and nonlinear steplike warming: the two simplest models of change. The results (undertaken in another paper) suggest that each side is both right and wrong. Warming is largely steplike, caused by interactions between climate change and variability confirming the stepladder–escalator model, but in a manner consistent with global warming theory. Over many decades, these steps integrate into a complex trend, so the critics are wrong in their interpretations of what these steps mean. Communicating complex and simple trends is more theoretically robust than communicating short- and long-term trends. If climate change follows a nonlinear pathway on decadal timescales, methods for detection and attribution, decadal prediction, the characterisation of risk and their communication all need to change. Scientific narratives need to move on from the existing gradualist narrative to those that addresses complexity and communicate uncertainty more effectively.

Introduction

The website Skeptical Science (2015) has a famous graph titled "How Contrarians View Global Warming", where annual global mean surface warming anomalies since 1970 follow a series of horizontal steps, like an escalator (Figure 1a). Another graph titled "How Realists View Global Warming" has a single, linear trend line through it (Figure 1b). These two graphs elegantly summarise a major debate in the climate wars. The Realist view is that long term, surface air temperature follows a trend consistent with the theory of greenhouse-gas-driven climate change, whereas Contrarians argue that deviations from that trend either disprove the theory or suggest that the risk is less than stated. However, when a rule-based statistical test developed to detect multiple step changes (Ricketts, 2015;Ricketts and Jones, 2016) is applied to those records, we get a result closer to the contrarian one (Figure 1c). Could the two arguments be both right – and wrong?

Skeptical Science (2015) summarise this debate as the contrarian camp "misunderstanding the difference between the long-term signal of climate change and the short-term noise of climate variability." While many contrarians clearly use this 'misunderstanding' to create doubt and delay, they may have one thing right. Does atmospheric warming on decadal timescales contain abrupt changes? If so, could those changes be more than just noise? In that case, both the Contrarian and Realist positions would be partly right – and partly wrong. Even if Figure 1a may be cherry-picked, Figure 1c is not and produces a similar outcome. If nonlinear changes are present in the climate signal, or even in the so-called noise, omitting them from the record will lead to climate-related risk being underdetermined. If risk is underdetermined, adaptation planning may fall short (Jones et al., 2013).

The focus on Skeptical Science's framing of what constitutes reasonable and unreasonable evidence for warming is not intended to be directly critical of them but to reflect a broad consensus held within the scientific community. Skeptical Science are just very good at communicating this. There are two arguments being made in this debate. One is that the climate change signal is best represented by the long-term trend and short-term noise deviating from that trend is variability (Risbey, 2015;Skeptical Science, 2015;Lewandowsky et al., 2016). The Contrarians have two main arguments that directly address this point. One is that if abrupt nonlinear changes like steps are detected in temperature it is not due to greenhouse gas-induced forcing, because by definition that is gradual and that other forcing mechanisms are involved

(Douglass and Knox, 2012;Douglass and Knox, 2015a;Tisdale, 2015). The other is that if the trend component of warming since 1998 has paused or stopped, then the science is overstated and warming is less risky than posed (McKitrick, 2015).
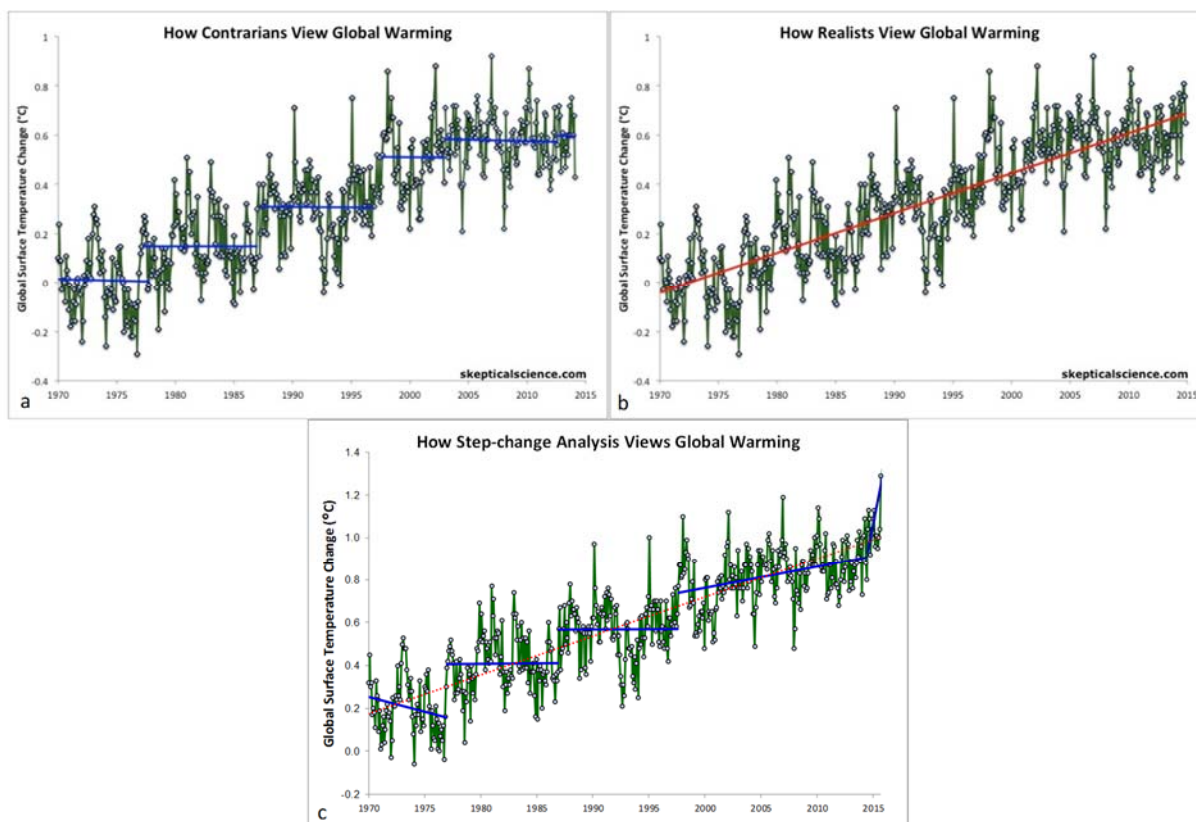

Figure 1. How step-change analysis views global warming compared to how a) contrarians and b) realists view global warming as argued and illustrated by Skeptical Science (2015), compared with c) the step and trend analysis based on the method described in Ricketts and Jones (2016). Monthly anomalies from the NASA GISS mean global warming to 2015 (a, b) and to March 2016 (c). Note that anomalies are 1951–80 in a) and b) and 1880–1899 (pre-industrial) in c). The time series are also slightly different due to ongoing adjustments to improve data quality.

Each of these arguments is based on statistical induction, where explanation goes beyond the data to suggest that if forcing is gradual and trend analysis can be used to explain the warming, then a gradual warming signal in response to greenhouse gas forcing is a theoretically held position. This position is being argued both pro and con by both sides, but with qualifications. The Realists argue that long-term trends are a valid measure of the climate change signal but that short-term trends are not. The Contrarians have two bites at the cherry: maintaining that a low or no trend is evidence of little or no warming and that an abrupt change is evidence of something else. Suggested alternatives include the El Niño–Southern Oscillation (ENSO; Tisdale, 2015) and solar influences (Douglass and Knox, 2015a, b), but no mechanism has ever been satisfactorily presented.

Statistical inference

All statistical models imply an underlying process, which can be explicit or implicit (von Storch, 1999;von Storch and Zwiers, 2001). In the absence of clear theory, a user of a statistical model will intuit a structural model of how the system works that makes internal sense to them. Benestad (2016) suggests the different models held by the research community vary widely – this variation has possibly hampered the construction of a clear scientific narrative for how the atmosphere warms. The Contrarian position is based on inductive

reasoning that treats linear trends as a proxy for theory: if observed warming does not strictly conform to a smooth trend, then the theory is either weakened or falsified.

By representing warming as a simple trend, it is unclear how much the Realist response in Figure 1b is reinforcing this view – most climatologists acknowledge the nonlinearity of the climate system, but the scientific community-of-practice lack a clear narrative as to how this manifests, beyond being random (e.g., Solomon et al., 2011;Deser et al., 2012). We consider the short- versus long-term trend argument to be scientifically weak – there is no objective rationale for deciding between what is short-term and what is long-term. It is a function of the statistical model applied and the question asked. This lack of objectivity is almost certainly being gamed. While the mainstream scientist wants to correct the record and counter misleading arguments, the critic is there to create a visible and audible opposition. It is important to the contrarian cause that that this opposition comes across as a scientific disagreement. This can often be achieved by using some scientific-sounding statistics and half-truths dressed up as common sense. It does not have to be correct and there is no penalty for the critic if they are shown to be wrong – as long as they live the day to repeat those claims elsewhere.

This paper draws on the philosophy of science to investigate this debate around trend analysis and to suggest ways forward for developing a stronger research programme for climate science:
1. It examines the basis for constructing sound scientific argument by describing scientific research programmes (Lakatos, 1970, 1978) and identifying the key core and auxiliary global warming theory.
2. It unpacks theory falsification and confirmation to identify spurious arguments aimed at discounting global warming theory based on dogmatic empiricism.
3. It develops a theoretical framework for severe testing (Mayo, 1996, 2005;Mayo and Spanos, 2010, 2011) capable of investigating the alternatives presented in Figure 1. Severe testing applies a framework for using known theory to bound plausible hypotheses, then develops severe tests using statistical models capable of selecting a single hypothesis from its alternatives.

A related paper looks at why linear trend analysis is the preferred model for analysing climate, reasons that are mainly historical and methodological rather than being based on theory (Jones, 2015b). This paper takes that argument further to develop a testing framework capable of reconciling the linear and nonlinear components of climate change.

Scientific research programmes

Here we explore what constitutes weak and strong scientific evidence within a scientific research programme to understand how global warming theory may be confirmed or rejected. It is a prelude to asking "What is the strongest scientific justification for global warming theory?"

For the interlinked biophysical processes that make up the climate system, the underpinning theory is complex, taking in different disciplines such as physics, climatology and other earth sciences, chemistry and biology. Lakatos (1970, 1971) describes a scientific research programme as having core and auxiliary theory. The core theory underpins the scientific understanding of a process, allowing statements to be made describing its fundamental behaviour. Auxiliary theory fills in the gaps and links to other areas of theory, but is not material to conclusions arrived at using core theory. The acceptance of core theory requires multiple lines of evidence that elevate that theory above its competitors. If there are rival core theories, they must all carry a degree of plausibility.

Of the core theory, Lakatos (1971) says it has "a conventionally accepted (and thus by provisional decision 'irrefutable') hard core, surrounded by a belt of auxiliary hypotheses". This addresses the issue of scientific consensus and is consistent with scientific paradigms (Kuhn, 1970). An up-to-date description for a complex biophysical system would be to have an interlocked network of theories, with a node forming the core theory for a set of central propositions, and auxiliary theory surrounding that node and linking it with others. Within earth system science, this network would include nodes for radiative forcing, terrestrial hydrology, convective

behaviour, and for the carbon cycle on land and in the ocean and so on. Each node would provide core theory for its central processes and auxiliary theory for other nodes. Some fundamental aspects, especially relating to the laws of thermodynamics, will contribute to core theory in a number of areas. These could be illustrated using network diagrams, showing the key nodes connecting with strong and weak links, depending on whether they are core or ancillary. If the focus of investigation changes, these links will reconfigure onto different core theory and links. For example, temperature and rainfall are related but investigating the core theory in terms of how they may respond to a change in external forcing has interrelated but different nodes.

Within a networked system of scientific system of understanding, some areas will be strongly represented by theory while others are weak. By matching these strengths and weaknesses to the available evidence we can survey the depth of our understanding of the physical processes within the climate system. This includes epistemological content – what justifies our knowledge as being scientific? Do we understand it from a theoretical perspective or has it been constructed by inference from method?

The conclusion from this section is that any scientific controversy needs to be addressed from the perspective of the core theory, even if that theory is difficult to communicate in a populist environment. Public awareness of core theory can be developed over time and it establishes a platform of scientific confidence from which uncertainties in related areas can be more readily addressed.

Scientific research programmes can be evaluated in terms of progressive and degenerating problem-shifts Lakatos (1971), so like much of philosophy of science, work best in hindsight (Curd and Cover, 1998;Mayo and Spanos, 2010). The puzzle solving of normal science as described by Kuhn (1996) will therefore largely concentrate on auxiliary rather than core theory. Instead of being a revolutionary process as described by (Kuhn, 1996), most changes in rival progressive and degenerating core theory follow a long, drawn-out process between the time a change is first proposed to when it is accepted (Lakatos, 1971;Laudan, 1977;Curd and Cover, 1998). This is what Lakatos (1971) means by provisionally irrefutable – overturning core theory is a substantial task, as described below.

Climate change as a scientific research programme

The successive reports of the Intergovernmental Panel on Climate Change (IPCC) and similar programmes auspiced by bodies such as World Meteorological Organisation and the World Climate Research Programme, applied to Lakatos' (1970, 1971) descriptions, serve as a mature scientific research programme. The understanding of climate change science has been improving over time, as widely documented by IPCC reports since 1990 (IPCC, 1990, 1996, 2001b, 2007, 2013) and the general science literature (Weart, 2008).

The core theory for the greenhouse effect is as sound as any in the earth sciences, its earliest version extending back almost 200 years (Weart, 2008;Pierrehumbert, 2011). If greenhouse gases are added to the atmosphere, the energy balance in the earth system will change, warming the climate. This core understanding of greenhouse theory is articulated in recent reviews:

1. Radiative transfer theory states that the accumulation of greenhouse gases in the atmosphere absorb longwave infrared radiation (LW) making the atmosphere warmer than it would otherwise be (Weart, 2008;Pierrehumbert, 2011). The planet absorbs shortwave radiation (SW) from the sun (amongst a broad energy spectrum of longwave to cosmic rays) and returns much of it to space as longwave radiation. Part of the heat transfer process occurs directly in the atmosphere and some is surface heat re-radiated into the atmosphere. Some SW radiation is reflected directly back into space by clouds and surface reflectivity but much of the shortwave absorbed by the surface (and a small amount within the atmosphere) is emitted as IR. If greenhouse gases increase, absorption and re-radiation of IR within the atmosphere ensures that more IR is directed downwards than upwards (Pierrehumbert, 2011;Trenberth, 2011). The limited space for IR absorption within spectral absorption bands of greenhouse gases also means that as saturation of those bands increases at lower altitudes, successively higher levels of the atmosphere no longer lose IR to space because they are taking up IR

    from below (Pierrehumbert, 2011). This requires radiation loss to occur higher in the atmosphere and has a fingerprint distinctly different to any other kind of forcing.

2. The reduction in the IR flux to space then creates an energy deficit at the top of the atmosphere (Weart, 2008). This deficit requires the planet to alter its energy flows by warming and modifying latent and sensible heat fluxes (Trenberth, 2011). Warming occurs through a combination of radiative and convective transfer, until there is a sufficient capacity for IR loss from the upper atmosphere sufficient to achieve planetary energy balance (Benestad, 2016). Warming will continue until outward heat flux from the atmosphere is equalised with energy coming in. Positive feedbacks within the atmosphere ensure that there is more warming than would be caused by simple 'black body' radiation (Stevens and Schwartz, 2012;Benestad, 2016).

The evidence underpinning this understanding is comprehensive. It includes observations measuring radiative absorption across the infrared spectrum, the energy deficit at the top of the atmosphere, the addition of energy to the Earth system, vertical and horizontal warming fingerprints and changes in a host of systems affected by higher temperatures. All attest to a warming planet (IPCC, 2013, 2014b). At one time atmospheric temperature would have been the key piece of evidence for this, but with the advent of space-based observation systems and energy budget measurements, this is no longer the case (Stevens and Schwartz, 2012;von Schuckmann et al., 2016).

There is insufficient room in this paper to take this topic to its full extent but the treatment of theory within research programs is hugely important and is not treated as systematically as it could be. Climate science exercises its scientific credentials though the Intergovernmental Panel on Climate Change (IPCC) which needs to apply a more rigorous approach to addressing theory as distinct from evidence (Ebi, 2011).

Scientific reasoning and falsification

The two main methods of scientific reasoning are deduction and induction. A deduction, usually a theoretical statement, has a conclusion based on the logical outcome of specific premises. This was once the ideal for science, in that science was thought of as deduction based on facts (Lakatos, 1978). Many people, believing a certain stereotype, still think this is the way science works. Deduction is now recognised as a type of scientific reasoning that is necessary but insufficient.

Induction is reasoning by inference that can be by analogy, statistical inference (often using probabilities) or induction to a particular (if all A have been B so far, then the next A will be B), to name three (Curd and Cover, 1998). Abduction, or inference to the best explanation to identify a 'true' theory from alternatives is a special case. It often prefers the simplest explanation as the best (parsimony). All the above methods of reasoning have been shown to be fallible, so a practical solution is to propose a hypothesis with true and false conditions, and to use different forms of reasoning to test the robustness and consistency of those conditions, while avoiding circular or magical thinking.

Contrarian claims are based on an idealised view of the scientific method that is very close to the deductive ideal above, combined with a simplified Popperian view of falsification, if one 'fact' is out of place, the theory is disproved (Coady and Corry, 2013;Mercer, 2016). The commonly-made assertion, that a single experiment ("one item of contrary evidence" (Plimer, 2009, p. 14;Zimmerman, 2009)) can falsify a system of theory (naïve falsification), is mistaken (Lakatos, 1970). Popper's name is invariably invoked to support naïve falsification (e.g., Symmons, 2011), but this misconstrues what he actually argued – the logic of falsification is simple but its practical application is difficult (Popper, 1992, 2002).

Falsification and confirmation comprise two classes of theory-statement that test consistency and inconsistency with a theoretical position (Thornton, 2015). For both falsification and confirmation, the truth

content should exceed the degree of falsity[1]. Even then, both theories may be false (Miller, 1974;Tichý, 1974). Furthermore, a theory is not abandoned until there is a better one to supplant it (Popper, 1992).

The claim that non-trend-like atmospheric warming disproves greenhouse theory comes from the proposition that gradual forcing should produce a like response in atmospheric temperature (Lloyd, 2012). This position is dogmatic empiricism – the insistence that theory should directly match observations. This claim is combined with naïve falsification, which contends that if observations do not conform directly to a theoretical proposition, then a theory is falsified. Put simply, if warming does not conform to a trend, global warming theory is either false (Carter, 2006;Plimer, 2009) or overstated (Carter et al., 2006;Douglass and Christy, 2009;Singer, 2011). This reasoning is based on the fallacy of dogmatic falsification – that a single 'anomalous' observation can disprove a theory.

This argument does not threaten core greenhouse theory. Because observations are theory-laden (Edwards, 1999;Lloyd, 2012), an observation or experiment needs to be linked to a theoretical proposition (van Fraassen, 2008). The direct linking of linear trends in atmospheric warming to gradual forcing without recourse to an explanation of cause and effect therefore fails. This reasoning also causes problems for the orthodox position (Mercer, 2016). Asserting that long-term trends represent the climate change signal whereas short-term trends do not, requires a theoretical basis for making such a distinction. We are unaware of any such distinction in the literature and believe it to be impossible, following Cohn and Lins (2005) and Koutsoyiannis (2010). If climate is the combination of quasi-oscillatory cycles on multiple timescales (years to millennia), the way these combine may be unpredictable, although they will conform to certain conditions, such as obeying power laws.

While long-term trends of 50+ years can safely be identified (e.g., Risbey, 2015), a precise distinction between what constitutes short and long-term is unattainable. Empirical studies, such as that by Santer et al. (2011) who nominate a minimum period for measuring human influence can (unwittingly) lead to those results being gamed. For example, Santer et al. (2011) identified that at least 17 years in lower tropospheric satellite observed and simulated temperature were required to identify human influences. However, that finding set an error bound, not a level of significance. In 2014, 17 years (inclusively) from 1998, these results were widely misused to suggest a significant threshold if scientific 'proof' had failed. For example, the former US Presidential Republican Party candidate Senator Ted Cruz repeated the "no warming for 17 years" claim, which then had to be refuted (Santer and Mears, 2015). Because the division between short and long-term cannot be closely identified it can therefore always be exploited.

The key experiments using the physics that constitute the core theory as described above show that natural and anthropogenic forcing produce distinctly different warming pathways over the 20th century (IPCC, 2001b, Figure SPM2-5; 2014a, Figure 1.10). Statistical models of how the atmosphere warms do not contribute to this core theory – rather the confirmation comes through the combination of modelling, radiation deficits at the top of the atmosphere and the accrual of energy in the Earth system (von Schuckmann et al., 2016). Any one of Figure 1a, b or c provides the same level of evidence because they analyse the same underlying data. If by this reasoning Figure 1a cannot be used to invalidate the theory, neither can Figure 1b be used as 'proof'.

Stronger confirmation of global warming can be obtained from increasing ocean heat content. About 93% of historically added heat currently resides in the ocean (Roemmich et al., 2015), thirty times that of the atmosphere. Between 1955 and 2010, the amount of heat added to the atmosphere was about 0.8×10^{22} Joules, compared with the 24.0×10^{22} Joules added to the top 2000 m of the ocean (Levitus et al., 2012). The

---

[1] Here true and false are not considered universal but local, given the best evidence at hand following Roush, S., 2005: *Tracking truth: knowledge, evidence, and science*. Clarendon Press, Oxford, 235 pp. This side-steps the debates between philosophical realists and non-realists as to whether a theory is (universally) True or (locally) true and is intended to frame a point where both realists and non-realists will agree in order to show the separation between science and pseudoscience.

amount of heat added to the atmosphere is roughly equivalent to the below-ground warming of the land surface (Beltrami et al., 2015).

Based on philosophical grounds, global warming theory cannot be refuted by any of the following claims:

- Failure of prediction is often invoked to state that climate model projections and observations of global mean temperature, historic and contemporary, are incompatible. This conflates conditional scientific prediction with unconditional prediction (Popper, 2002). Recent work shows the opposite is the case: the more comprehensive the model input and processes, the better they match observations (IPCC, 2013). Hindcasts with natural and anthropogenic forcing show that the physical models pass a severe test (IPCC, 2001a, Figure SPM2-5;2014a, Figure 1.10).
- Failure of the atmosphere to warm at a constant or sufficient rate, as a manifestation of a theoretical proposition. This is the fallacy of dogmatic falsification (dogmatic empiricism + naïve falsification), where the results of a statistical analysis are held to represent a theoretical position that can be falsified by observation.
- Conflation of uncertainty or lack of knowledge with a degree of falsification or empirical failure. Being uncertain is not the same thing as being wrong and being certain (dogmatically so) is not the same thing as being right.
- The 'observations are real' and 'models are artificial' proposition that overlooks the fact that observations are theory-laden (Kuhn, 1996) and can only be considered objective within the appropriate context (Longino, 1990), the same conditions that models operate under.
- Falsification without proposing an alternative hypothesis/theory with its own set of potential verifiers and falsifiers (Popper, 1992) and justification as to why .

The only negative cases that could threaten global warming theory would either have to show why future climate would behave differently to our current understanding of how the past has behaved, corroborated by physical models and observations; or that climate is being affected by drivers we are unaware of (e.g., Sloan and Wolfendale, 2013). This would need to be supported by a credible experimental program detailing this different behaviour and how it may play out in future.

Statistical induction and severe testing

The type of reasoning behind Figures 1a–c and similar analyses designed to separate signal from noise is statistical induction. Mayo and Spanos (2011) advise care in distinguishing between the error statistic and the probability of confirmation – likelihood tests passing criteria such as $p(H_0)<0.05$ as the error statistic run the risk of being interpreted as addressing scientific hypotheses with the same level of confidence. A common and legitimate usage of these two is where error statistics are used to estimate a significant trend, then that trend is confirmed as an important phenomenon through induction to a particular; e.g., where sufficient members of an ensemble of climate models meet the same criteria.

Parsimony is used to justify linear trend analysis as the simplest explanation in the light of unresolved uncertainty (i.e., inference to the best explanation). This practice is criticised on the basis that using parsimony and likelihood tests to select a structural model are inferior to tests that apply experimental reasoning and that are statistically adequate for the problem in question (Mayo, 1996;Mayo and Spanos, 2010;Spanos, 2010b). Mayo and Spanos argue that conditions for severe testing should be probative, rather than probabilistic. If test *T* has no likelihood of finding flaws in *H*, then it is not a good test. Mayo (1996) calls this the fallacy of acceptance: no evidence against the null is interpreted as evidence for it, and evidence against the null is interpreted as evidence for an alternative.

Inference to the best explanation can be rescued if it passes the test: "no available competing hypothesis explains a fact as well as *H* does" (Musgrave, 2010). Mayo (2005) provides criteria for severe testing that is even stricter: "Data *x* in test *T* provide good evidence for inferring *H* to the extent that hypothesis *H* has passed a severe test with *x*". This provides a level of rigour beyond the experience that might be brought to bear by an individual experimentalist's preferences. The problem with using cognitive values (e.g., parsimony, past

success) to decide the 'best' explanation is that such values are may be historical and methodological rather than being probative, as has been the case for climate (Jones, 2015b).

Development of criteria for severe testing requires aligning structural models that define the problem in theoretical terms with experimental and statistical models to be used for testing. Structural models representing one or more hypotheses are formulated. Statistical models are then selected for their ability to represent these hypotheses, or aspects of them. To avoid under-determination or misdiagnosis, the underlying assumptions of applying data *x* to *H*, the (scientific) hypothesis associated with the structural model needs to be properly aligned with the (statistical) hypothesis attached to the statistical model being used to test it.

This approach eschews the behavioural or mechanical reasoning that suggests a particular action needs to be taken when a *p* value reaches a particular threshold. Cox and Mayo (2010) distinguish between probabilistic and behavioural reasoning where distinguishing between *H* and not-*H* needs to pass a severe test. Gigerenzer and Marewski (2015) refer to the automatically inferred conclusion based on a p-value threshold as mechanical reasoning and are just as critical. Understanding statistical error through *p* values is important, but determining how such probabilities should be calculated, and what levels clearly distinguish *H* from not-*H* is part of the experimental design and learning by experimental error (Mayo, 1996). Doing this requires a clear understanding between two different types of hypotheses: the scientific (*H*) and the statistical hypothesis (*h*), and linking them through highly probed, rather than highly probable, testing (Mayo, 2005).

The failure to adequately distinguish between scientific and statistical hypotheses within an environment of severe testing has fuelled much of the science wars. This failure comes from both sides of the debate, where the mechanical use of *p* values is used to support or debunk a particular proposition. For example, Mercer (2016) describes how both pro and con arguments utilise simplified versions of Popper's philosophy to justify their positions, often on the back of a statistical test.

In this light, the hindcasts of natural and anthropogenic drivers and their effect on global warming of the 20th century described in the previous section, provide an example of severe testing because they examine both *H* and -*H*. In the one example in the literature, Katzav (2013) asks whether the IPCC Fourth Assessment Report (IPCC, 2007) statement that most of the post-1950 warming is due to anthropogenic greenhouse gas forcing (high confidence >90%; Hegerl et al., 2007), passes a severe test and decides it does not. That conclusion draws on the above experiment (but not solely). We disagree with this conclusion for several reasons.

Katzav (2013) based his assessment on trend analysis, classifying any deviation from the simple trend as climate variability and thus noise, speculating whether it could be >50% of the total measured warming. He also lists variables not included in the climate model experiments (e.g., cosmic rays, land-use, biogeochemistry) as potential contributors to not-*H*, but does not test those assertions. For example, he cites Semenov et al. (2010) who assess the influence of North Atlantic–Arctic decadal variability on northern hemisphere surface temperatures, yet Semenov et al. (2010) cite Hegerl et al. (2007) to support their study. He also dismisses the energy balance argument too readily. This invalidates Mayo's (1996) requirement that both *H* and not-*H* be highly probed. Katzav's (2013) challenge cannot be resolved by simple inference, it needs to be severely tested, so he fails to show that Hegerl et al.'s (2007) confidence level of highly likely is misplaced.

The trend model in Figure 1b itself that Katzav (2013) relies upon has not been severely tested. Using error measures and goodness-of-fit statistics no one statistical model dominates, and several are plausible (Seidel and Lanzante, 2004;Menne, 2006;Trenberth, 2015). Trend analysis is actually the weakest of these statistical models because trend significance is the key measure – gradual behaviour is not an explicit measure of the test so is only implied. Slingo (2013) emphasises that no single statistical model allows the identification of a true model and that it would require clear identification of the physics involved to do so.

For Figure 1b, mental models supporting gradual warming would require a model of in-situ atmospheric warming supplemented by the gradual release of heat from the ocean, obscured by climate variability. We know of no theory that directly supports this model, so assume it is largely empirical or is the historical legacy

of simple models justified by parsimony (Jones, 2015b). Such models have played an important pedagogical role in understanding and explaining warming (e.g., Benestad, 2016) so are tightly linked to cognitive values (Jones, 2015b).

Theoretical considerations

The climate of the ocean-atmosphere system is largely a combination of two physical subsystems that have different core theory and so exhibit different behaviour. The first is the radiative system in the atmosphere, where radiative conversions to other forms of energy such as infrared are considered to be additive (Ozawa et al., 2003). The second is the thermodynamic system in the atmosphere and ocean, which is characterised by turbulent flow, where heat energy is converted into potential energy, the source for kinetic energy. Internally, the climate system is always out of equilibrium, so this kinetic energy transports heat from the equator to the poles. Based on work by Lorenz (1963, 1975), Hasselmann (1976, 1979) and others (Palmer, 1993, 1998;Palmer, 1999;Ozawa et al., 2003;Lucarini et al., 2010;Lucarini and Ragone, 2011), these processes are understood to be fundamentally nonlinear.

Ozawa et al. (2003) argue that the background radiative down-conversion of energy from solar to terrestrial temperatures is essentially a linear process, which is independent of the process of entropy production related to nonlinear turbulence. This argument has strong implications for how heat energy is produced and distributed within the climate system. Ozawa et al. (2003) view the radiation system as an energy source, but see the turbulent dissipation of that energy as entropic. As such, it would be entrained within the quasi-oscillatory (Lorenzian) modes of what is considered natural climate variability, perhaps modifying that variability in the process, moving the most amount of energy with the least amount of work.

If we accept this argument, then a small addition of energy to the climate system will increase disequilibrium between the surface and the top of atmosphere and also the equator and the poles, leading to increased heat transport in turbulent media – the ocean and the atmosphere – both horizontally and vertically. Heat production through the trapping of heat by greenhouse gases would therefore be linear and its transport within the climate system, nonlinear.

Lorenz (1975) identified two kinds of predictability that can be linked to these processes:
- The first is due to changing boundary conditions with fixed initial conditions and is associated with long-term (multi-decadal to centennial) climate predictability (Lorenz, 1975;Hasselmann, 2002;Collins et al., 2011), and;
- The second is due to initial conditions with fixed boundary conditions and is associated with weather and shorter term (interannual to decadal) climate predictability (Meehl et al., 2013).

The first process will be quasi-linear in line with changing radiative forcing defining the boundary conditions, unless a singularity occurs, resulting in an abrupt positive or negative feedback (e.g., ice sheet collapse). The second kind of predictability is related to entropy, where the ocean-atmosphere system acts as a large heat pump, transferring moist static energy into work through irreversible processes. Heat is transferred from the equator to the poles via both the atmosphere and ocean, which exhibits self-organised criticality (Ozawa et al., 2003). Climate will seek a stationary state until such time as energy stipulates that another state can transport energy poleward and skyward more efficiently – this can be thought as when one state becomes unlikely (or approaches critical limits), it will transform into another that is more stable (Tsonis and Swanson, 2011;Tsonis and Swanson, 2012).

Metastable and quasi-periodic climate regimes, punctuated by rapid changes on timescales ranging from decades to millennia, are typical aspects of the climate system. Decadal scale oscillators that behave in this way include the Pacific Decadal Oscillation (PDO)/Interdecadal Pacific Oscillation (IPO) and the Atlantic Multidecadal Oscillation (AMO) (Mantua et al., 1997;Hare and Mantua, 2000;Mantua and Hare, 2002;Newman et al., 2003;Enfield and Cid-Serrano, 2006;Zhang and Delworth, 2007). The mechanisms behind some of these shifts are beginning to be understood. Linked to an abrupt shift between El Niño–Southern Oscillation (ENSO)

regimes in 1978, O'Kane et al. (2014) show that fast (weather) and slow build-ups of density anomalies combine to disrupt the thermocline in the equatorial Pacific. This is linked to a contemporaneous step change in global temperature (Jones and Ricketts, 2015).

Step changes in global warming have been deemed implausible because the year-to-year changes in radiative forcing are too small to contribute to such large changes (e.g., Cahill et al., 2015). However, this view overlooks ocean-atmosphere interactions. The dominance of recently accumulated heat stored in the ocean with respect to the atmosphere (a 30:1 ratio), means that a substantial redistribution of heat in the shallow ocean associated with a regime change could easily be transmitted to the atmosphere. The amounts of heat energy required would only be small part of the annual ocean-to-atmosphere heat flux. The added heat fluxes due to anthropogenic greenhouse gases are around 1% of the annual total flux (IPCC, 2001b).

Adopting an energy-balance approach, as a proportion of energy measured at the top of the atmosphere over the annual cycle that is absorbed within the earth system almost all is taken up by the oceans; storage on land is negligible (Fasullo and Trenberth, 2008b, a). From this stored energy, the ocean to land flux is substantial – much of which is transported as atmospheric moisture (Fasullo and Trenberth, 2008a). The land can heat substantially but will cool just as readily. This creates a straightforward either/or for radiative forcing. If the heat trapped by increased forcing is absorbed as part of the normal seasonal energy flux into the oceans, it will combine with processes that govern climate variability. If it remains separate as added heat in the atmosphere, it needs to be considered as an exception to the above process and a reason for that exception needs to be identified.

A number of model studies show increases in warming over land are due to warming in the oceans and subsequent transport of heat over land rather than through direct radiative forcing (Compo and Sardeshmukh, 2009;Lambert et al., 2011;Geoffroy et al., 2015). If ocean warming due to radiative forcing is compared to direct increases in sea surface temperature, the warming over land is little different (Dommenget, 2009). This ocean–land effect is much greater than the land–ocean effect, where increasing temperature over land has little effect on the ocean (Dommenget, 2009;Lambert et al., 2011). The oceans are estimated to contribute 80–90% of the warming on land in one estimate (Dommenget, 2009) with horizontal energy transport contributing 70% in another (Geoffroy et al., 2015). Uncoupling transport between land and ocean leads to little warming for either, showing it is the coupled relationship between the two that is important (Lambert et al., 2011).

If ocean surface warming is gradual, the land response will be gradual, but if warming arises out of nonlinear interactions between the ocean and atmosphere (involving land as per the coupled transport process above) nonlinear warming on land would follow almost immediately, perhaps driven by the processes described by Reid and Beaugrand (2012). Under this hypothesis, decadal climate regime change arising out of ocean-atmosphere interactions is capable of producing step changes in mean sea surface temperature. Consequently, these step changes could be transmitted to adjacent continental areas and/or though teleconnections. The atmosphere has no heat storage memory but the ocean does. If the added heat energy trapped by anthropogenic greenhouse gases follows the same path as natural heat energy trapped by naturally occurring greenhouse gases, the interaction of external forcing with internal variability is a logical result.

Warming hypotheses

Since the introduction into climatology of theory describing non-linear behaviour (Lorenz, 1963;Hasselmann, 1976), two separate paths of research have emerged. A theory-dominated path investigates nonlinear behaviour in the climate system through theory, analysis of model behaviour and the statistical analysis of break points and regime changes in observational data. A largely method-driven path covers the analysis of model and observational data and its use in detection and attribution, forecasting and projection.

These investigations have produced two main hypotheses that describe the interactions between climate change and variability over decadal timescales (Corti et al., 1999;Hasselmann, 2002;Branstator and Selten, 2009):

H1. Externally-forced climate change and internally-generated natural variability change independently of each other.

H2. They interact, where patterns of the response project principally onto modes of climate variability (Corti et al., 1999) or form a two-way relationship (Branstator and Selten, 2009).

These hypotheses suggest different structural and statistical models. *H1* maps onto the statistical model shown in Figure 1b. If decadal regimes were also to be considered to be moderating the rate of external warming; i.e., via gradual release of heat from the ocean, faster and slower rates of release may combine with a gradual rate of atmospheric warming to produce a segmented trend. *H2* would entail some form of nonlinear signal over decadal timescales, where warming interacts with processes governing decadal regimes, perhaps steps, steps and trends, segmented trends or a combination of these.

In the literature, the theory and method-driven paths are largely addressed separately (Jones, 2015b). The method-driven path giving rise to *H1* dominates how climate information is analysed and communicated. This informs the gradualist narrative of adaptation, which leads to adaptation planning focusing on a gradually changing mean climate punctuated by random shocks due to climate variability (Jones et al., 2013). When theory is discussed, however, the possibility that climate change and variability may interact is also considered plausible (Branstator and Selten, 2009;Solomon et al., 2011;Kirtman et al., 2013).

The two hypotheses, *H1* and *H2*, imply at least two different structural models. Theoretical understanding of the underlying processes can be used to set up structural models for severe testing using statistical hypotheses.

For *H1*, close adherence to a warming trend implies that gradual warming occurs in the atmosphere. If so, this must occur via either of the following two processes (1, 2) or a combination (3):

1. Radiatively-forced warming occurs in the atmosphere and the heat generated is retained there (*in situ* warming). Statistically, this would manifest as gradually increasing temperatures, especially over land. It would also imply a trend in lower troposphere satellite temperatures as the whole airmass warms.

2. Most of the heat generated by added greenhouse gas forcing goes into the ocean and is re-released gradually into the atmosphere. Again, this would imply trend-like warming, especially over the oceans, with the land following suit, but with some variation if decadal changes in shallow and deep-ocean mixing of heat are taken into account. Discussions in the literature are not clear as to whether the oceanic component would be due to varied take-up or release of heat from the ocean, but most discussions imply release as do simple models (Raper et al., 2001;Wigley, 2008).

3. If both 1 and 2 are operating, then the warming rate in the atmospheric component would be monotonic (proportional to forcing) and the contribution from the ocean governed by interannual and decadal variability. This would be best represented by a segmented trend if decadal-scale deep and shallow ocean mixing of heat is a key factor.

In situ atmospheric warming would require gradual trend-like behaviour to be observed at regional to global scales. If heat is taken up then emitted gradually from the ocean, segmented trends could be produced if the rate of release is modified by alternating periods of deep, normal and shallow ocean mixing. Given the large thermal mass of the ocean, and the relatively small amounts of warming required to heat the atmosphere (Fasullo and Trenberth, 2008a), relatively small regions of release may be possible (i.e., as occurs during El Niño episodes). If gradual warming is a mix of *in situ* warming and ocean release, then combinations of trend and segmented trend would be possible. This is the dominant paradigm, but generally not stated clearly as such in the literature – it is received wisdom and explained as "this is the way things are".

For *H2*, interactions between change and variability on decadal timescales could see change interacting with decadal regimes with step-like responses as in Figure 1c. A plausible mechanism is that the ocean absorbs

most of the heat generated in the atmosphere with minimal heat remaining in situ as described above. The ocean absorbs heat until it becomes unstable, and acting like a heat pump, releases some of this as part of a regime shift.

The main areas of core science discussed earlier reflect the trapping of heat by greenhouse gases on the one hand, and its measurement as a temperature change on the other – positive feedbacks are part of the temperature response. What is not known is what happens to the heat between the time it is trapped in the atmosphere and measured as a temperature response (Figure 2). The presence of step changes in climate model temperature output suggests nonlinear behaviour, but none of the diagnostic tools currently in use assess this type of behaviour, preferring to linearise the output in order to study it. Measures like effective forcing that factor in net ocean heat uptake do not consider the physical process explicitly, but assess net uptake as a factor of the surface temperature anomaly (e.g., Winton et al., 2010). In this vein, Held (2005) discusses the need to need to close the gap between simulation and understanding in climate modelling, nominating nonlinear behaviour as a particular area of concern.
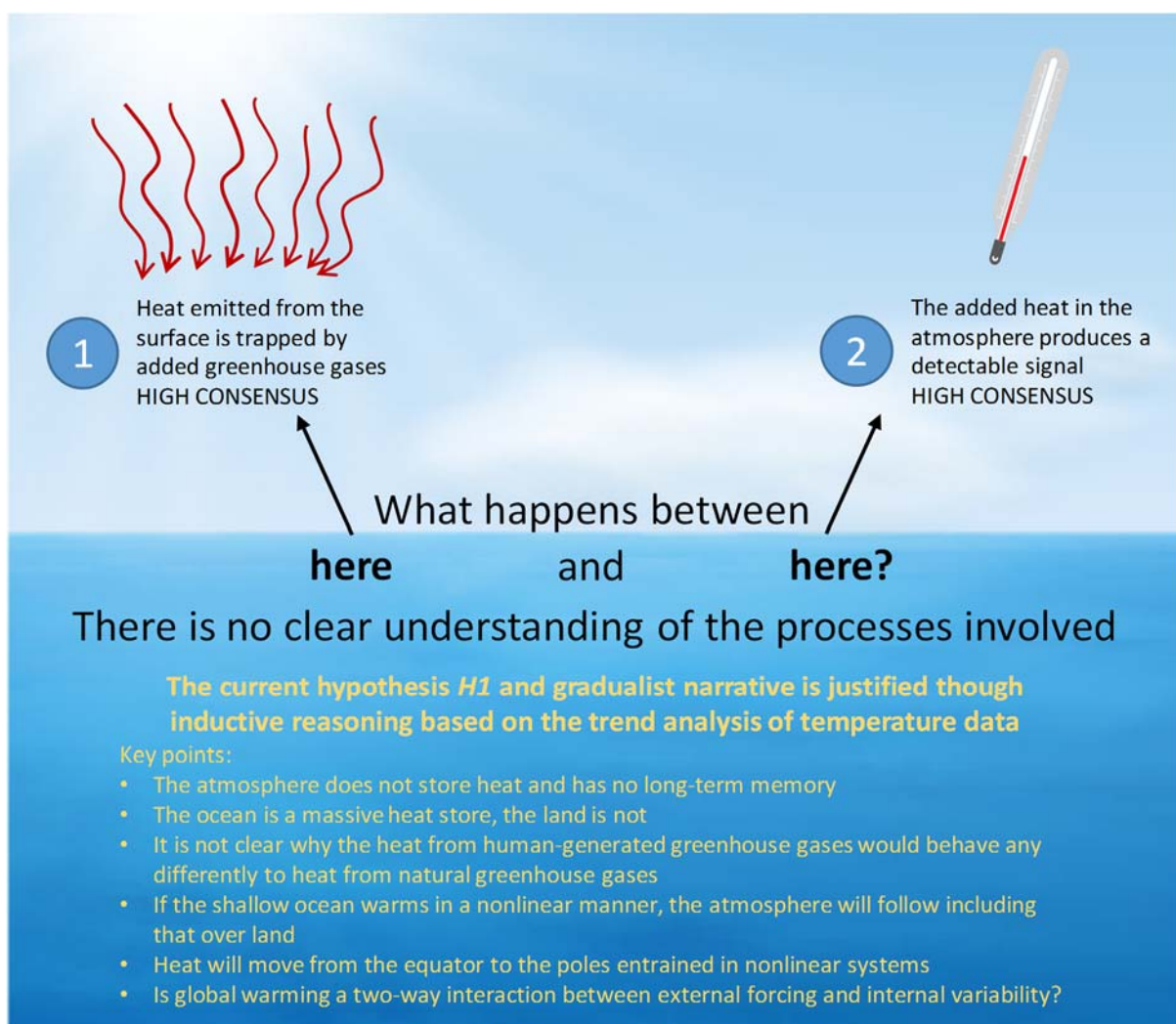

Figure 2: Infographic cartoon showing two areas of core theory where there is high scientific consensus questioning the process that takes place in between the time that heat is trapped in the atmosphere and when it is detected as warming.

If the atmosphere is unable to store heat on its own, temperature change will depend on the supply of energy from an active store, all other things being equal (i.e., solar inputs, cloud behaviour, land surface characteristics), which is the ocean (land is a passive, low-exchange store). Stable regimes being maintained during periods of increased forcing would imply the added heat is entering the ocean. Raper et al. (2002) show

that models with larger transient warming have a higher ocean heat uptake. Winton et al. (2010) show this process also happens under transient warming, heat uptake increasing with temperature over time.

We posit that there are two types of regime change distinguished by the relationship between entropy sources and sinks. One type maintains the overall energy balance at roughly the same level, but reorganises the major pathways that energy takes between source and sink when regional inequilibria become critical. This type of regime change is that typically observed with the reorganisation of decadal oscillations such as the PDO or IPO in a stable climate. An example is the late 1940s shift in the PDO/IPO, which is associated with a change between drought- and flood-dominated regimes in eastern Australia (Warner, 1995), which registers as a step change in rainfall (Vivès and Jones, 2005). However, the relationship between maximum temperature (*Tmax*) and rainfall (*P*), correlated variables in the region that serve as a crude measure of the regional latent and sensible heat balance, does not change. As *P* goes up, *Tmax* goes down, so the overall energy balance is maintained (Nicholls et al., 2004). Key correlated variables within the system undergo a shift but the relationship remains statistically stationary (Jones, 2012).

The second type is in response to external forcing that pushes the system out of equilibrium (positive) or relaxes it (negative). In this case, the system will either increase or decrease the rate at which it transports energy horizontally and vertically. An externally-forced regime shift would involve positive/negative step changes in sea surface and air temperatures coinciding with identifiable changes in ocean-atmosphere regimes as seen in ordinary regime change. With respect to attribution, correlated variables on adjacent landmasses reflecting sensible-latent heat energy balance such as *Tmax*/*P* and *Tmin*/*Tmax*, will see the first variable undergo a step change with respect to the second variable, denoting a non-stationary change in the relationship because of added warming/cooling. Consistent with the modelling studies in the previous section step changes in sea surface temperatures accompanying regime change would spread to the land surface within one to two years.

In this model, the ocean behaves homeostatically. Under positive forcing it will absorb increasing amounts of heat, but surface temperatures will remain stable until a critical state is achieved. A regime shift to a warmer state occurs, releasing heat into the atmosphere as it does so. Over the short-term, the warming process would be stochastic and over the long-term, quasi-linear, producing a complex trend. Negative forcing from aerosols, volcanoes or solar minima would delay regime shifts, or if large enough, produce downward steps in temperature.

It is generally understood that the atmosphere does not distinguish between natural and anthropogenic CO_2 when trapping heat, so it is not clear why the heat that is trapped would separate into natural and anthropogenic components and not become entrained in the normal processes governing climate variability. The next section looks at answering this question by using severe testing to distinguish between trend-like and step-like change on decadal timescales.

Severe testing

Understanding how the climate warms over decadal timescales is an inverse, ill-posed mathematical problem, similar to that for borehole (Şerban and Jacobsen, 2001) or satellite (Aires et al., 2001) temperatures, but is generally not considered as such. Atmospheric temperature is influenced by processes that are the product of external forcing and internal variability subject to some undetermined delay. Global mean temperature is also the integration of regional signals, so is spatially complex. Understanding the nature of these responses on multiple scales is important for characterising climate risk.

A hypothesis *H* passes a severe test *T* with data *x* if (Mayo and Spanos, 2010):
1. *x* agrees with *H* and,
2. with very high probability, test *T* would have produced a result that accords less well with *H* than does *x*, if *H* were false or incorrect.

Here, we consider the substantive null of model adequacy approach described by Mayo and Cox (2010) – specific results are used to provide theoretical evidence that can relate statistical tests with a scientific null H_0. Rather than H and H_0, rival hypotheses $H1$ and $H2$ are represented by statistically distinct models. This requires developing structural models representing scientific hypotheses $H1$ and $H2$, as detailed in the previous section. Statistical tests and hypotheses ($h1$, $h2$) then need to be developed from the structural models. The latter are subject to Type I and Type II errors. Sometimes such errors are used to justify scientific hypotheses but because one is a statistical error and the other is probative, they should be kept distinct (Spanos, 2010a).

Scientific hypotheses need to be examined closely in order to identify the probative conditions required to be built into one or more statistical tests for severe testing. These are then subject to error-based statistical testing to ensure that both H and $-H$ and h and $-h$ have sound and testable bases:

- $H1$ hypothesis – The climate signal is gradual and follows a (probably monotonic) trend and variability is noise. Non-linear changes in climate data are due to internal climate variability and are random. The hypothesis is confirmed once a trend is judged as significant.
- $H1$ null hypothesis – a trend has not emerged from the variability.
- $H2$ hypothesis – externally forced and internally generated climate processes interact with each other producing a nonlinear climate signal on decadal timescales.
- $H2$ null hypothesis – the climate signal is gradual and any nonlinearity is of internal origin ($H1$).

This two-by-two structure is asymmetric and points to the fact that $H1$ has never really been severely tested – the usual $H1$ null hypothesis tests whether a trend has emerged beyond a given level of significance. Given that trend analysis in itself cannot confirm whether a trend is simple or complex, other tests, such as change-point analysis are needed to identify whether nonlinear behaviour is present.

Change points in climate data are increasingly being located by a variety of different techniques (Rodionov, 2005;Reeves et al., 2007;Overland et al., 2008;Beaulieu et al., 2012;Fischer et al., 2012;Jones, 2012;Ruggieri, 2013;Cahill et al., 2015). The challenge for severe testing is to create tests that can distinguish between $H1$, -$H1$, $H2$ and -$H2$ with sufficient confidence to select one over the other. As emphasised by Mayo and Spanos (2010), likelihood criteria based on goodness of fit alone are unlikely to be sufficient to make a clear distinction. For example, for the quarterly anomalies of the timeseries shown in Figure 1c, the residual sum of squares from the linear trend, steps and step and trend models are 2.6, 2.4 and 2.1, respectively. These differences are insufficient to distinguish between $H1$ and $H2$, if $H1$ is represented by a trend (h_{trend}) and $H2$ by steps (h_{step}).

Six tests have been developed to analyse steplike behaviour and severely test step- and trend-like structures in temperature data:

1. Stratified analysis of change points: the timing and distribution of change points and their relationship with known regime changes and with each other. Change points aligning with known nonlinear processes indicate a causal link.
2. Identification of similar patterns of steps between observations and physical models indicates a physically coherent origin, rather than random stochasticity.
3. Using internal trends and shifts (steps minus internal trends) to estimate the gradual and rapid warming components in a record, and testing each of these against criteria such as total warming and equilibrium climate sensitivity (ECS) in observations and models separately.
4. Tests to diagnose stationarity – using a simple linear inverse model to measure the emergence of an anomalous signal from the background noise of variability, and whether it is gradual or steplike.
5. Testing of other variables including rainfall, sea surface temperatures, sea level rise, and air pressure, to see whether they undergo similar changes.
6. Statistical models applied to test underlying step- and trend-like structures in the data.

We have applied these tests in another paper (Jones and Ricketts, 2016) and conclude that the step model outperforms the trend model such that $pH2>>pH1$ (see Box 1). Building on the core theory and gaps shown in

Figure 2, we speculate that the additional warming produced by human-induced greenhouse gases are routinely absorbed into the ocean and well mixed, producing no appreciable warming at the surface. Growing instability due to accumulating heat eventually leads to a critical point being reached and regime change occurs, causing step changes in sea surface temperatures and the release of appreciable amounts of heat into the atmosphere causing step-like warming. This store and release mechanism could then be seen as behaving like a heat engine, under external forcing undergoing a series of step-like changes that conform to a long-term complex trend (Jones and Ricketts, 2016).

Box 1. Results of severe testing of h_{trend} (H1) and h_{step} (H2) summarised from Jones and Ricketts (2016)

1. **Stratified analysis of change points**
   - Global and regional analyses of steps show a highly coherent pattern of change points.
   - Shifts dominate total historical warming when internal trends between steps are subtracted.
   - Step-like warming in the latter 20th century aligns with known changes in decadal regimes.
   - Analysis of steps, internal trends and shifts in observations attributes higher proportions of warming to shifts at zonal scales (up to 100%), moving to lower proportions at the global scale.
   - This effect is larger in the mid-latitude regions and with sea surface temperature.
   - Surface and satellite temperatures undergo contemporaneous shifts at the global scale.
2. **Similar patterns of change in observations and physical models**
   - Correlations between step change frequency in the observed 45-member group of global and regional data and the CMIP3 and CMIP5 model runs analysed (1880–2005), are 0.32 and 0.34, rising to 0.78 after 2050 if six events (1963/64, 1968–70, 1976/77, 1979/80, 1987/88 and 1996–98) are grouped and other years analysed individually.
   - Fifty-eight members of a 107-member model ensemble (CMIP5 RCP4.5) show a step change in 1996–98.
3. **Nonlinear components of warming carry more of the signal than linear components**
   - Steps and shifts and internal trends show r^2 values of 0.87, 0.43 and 0.13, respectively, with simulated historical warming 1861–2005 in a 107-member ensemble (CMIP5 RCP4.5).
   - Steps and shifts and internal trends show r^2 values of 0.96, 0.54 and 0.59, respectively, with simulated total warming 2006–2095 (CMIP5 RCP4.5). For ECS, the r^2 values are 0.65, 0.52 and 0.18, respectively.
4. **Stationary and non-stationary periods are separated by step changes**
   - In three regions on three continents tested, and six climate model runs for SE Australia, the separation between stationarity and non-stationarity is steplike not gradual.
5. **Other variables show similar step changes**
   - Step changes exhibiting similar timing have been shown for tide gauge observations, rainfall, ocean heat content, forest fire danger index and a range of other climate variables.
6. **The best representations of underlying step- and trend-like structures in the data.**
   - For observations and selected model data the simple step-ladder model performs substantially better than the trend model for goodness of fit, residual sum of squares, cumulative residuals, cumulative residuals squared, White's test for heteroscedasticity, and moving 40-year windows of autocorrelation and White's tests.

Discussion

As the mainstay of statistical climatology, trend analysis and related methods have been extremely successful as measured by the cognitive values of accuracy, consistency, scope, simplicity and fruitfulness (c.f., Kuhn, 1977). Trend analysis' supporting scientific values date back to the scientific enlightenment, starting off as cosmological values (order, balance and reason), these values over time becoming cognitive values (Jones, 2015b). Despite the advent of nonlinear theory and the recognition that climate is a complex system,

climatology has retained the classical statistical model of trend analysis, applying it within a signal-to-noise structure. By not adapting this model to account for complex system behaviour, climatology is overlooking a range of key phenomena intrinsic to understanding how the climate system works, as is the case for economics (Spanos, 2010a;Arthur, 2015).

Contrarian claims about global warming having stopped or slowed are largely based on the inference that the use of simple trends serve as a theoretical statement about how the climate changes. The science community has been loath to challenge this framing directly because they lack a solid evidence-based narrative to counter it. The argument that the atmosphere carries roughly 3% of the total added heat so is only a small part of total global warming is often made and though correct, is easily side-stepped. Surface temperature is the most visible of climate variables and the simplest to analyse, so will be the focus of any attacks on the underlying theory.

The Realist response as articulated by Skeptical Science (2015), is that long-term trends are relevant but shorter-term fluctuations are not (Risbey, 2015;Lewandowsky et al., 2016). This justification is scientifically weak because the separation of what is short-term from long-term has no theoretical underpinning and cannot be clearly defined due to the complex nature of the underlying data (Koutsoyiannis, 2010). Empirical measures using error-based probabilities from simple trend analysis in the form of $pH_0 < 0.05$ are limited by the statistical assumptions of the method used.

A number of critics are aware of this point (McKitrick, 2014;Legates et al., 2015), which also has wide support within the hydrometeorological literature (Cohn and Lins, 2005;Koutsoyiannis, 2010). Step changes described within the conventional scientific literature are generally attributed to climate variability (e.g., Tsonis and Swanson, 2012), although it remains an open question (Meehl et al., 2009;Solomon et al., 2011;Chikamoto et al., 2012). A growing number of studies are describing such events as contributing directly to the change process (Hope et al., 2010;Reid and Beaugrand, 2012;Belolipetsky, 2014;Bartsev et al., 2016). This is especially true for hydrometeorological studies, which have a higher expectation of nonlinear response than temperature-based studies. Some critics continue to frame global warming theory as being synonymous with gradual warming, arguing that if nonlinear change is present it must have a different cause (e.g., Douglass and Knox, 2015a;Gervais, 2016).

The criticisms in this paper are not anti-trend. Trend analysis is the major tool in the climatologist's toolbox and will remain so. However, a much more robust way of analysing and communicating trends is to separate them into simple and complex trends. Complex trends are theoretically robust and simple trends a statistical tool. The proposition that radiatively forced climate will adhere to a complex trend over long (50+ years) timescales, has both theoretical credibility and is preferable for the characterisation of risk. Simple trend analysis will remain as a widespread and useful tool but should not be seen as representing a theoretical position, because it is only one part of a more complex process.

Any focus on atmospheric temperature needs to be addressed with precision. Atmospheric temperature is a key variable for the detection and attribution of change, but as we have emphasised above, is not an important contributor to core theory. However, it is a key determinant of climate-related risk. For that reason, its prominence will continue as will criticism of the associated science. However, each aspect of theory, detection and attribution, forecasting and characterisation of risk should be dealt with separately, using fit-for-purpose methods and tools (Jones, 2015a). These different aspects need to be recognised internally by the science community and these differences communicated more broadly. Atmospheric temperature is critical for determining risk, but not for addressing core climate theory. Simple trends are inadequate for both.

With respect to theory, stronger options for defending climate science can be based on communicating the core theory, identifying claims that violate philosophy of science principles and severely testing hypotheses that bridge key gaps in knowledge. Although these points are more complex to communicate than simple statistics, such a strategy could address the following issues:

- The core science is addressed directly rather than responding directly to external claims that have little or no scientific basis in themselves; i.e., work from a position of strength.
- It focuses on sound scientific practice underpinned by philosophical principles; e.g., core theory, passing severe tests, justified evidence. This is a useful pedagogical tool for teaching science.
- Failing to directly engage with nonlinearities by waving them away as random fluctuations and thus allowing them to continue to be used to challenge theory is a poor strategy.
- The use of simple and complex trends as a terminology instead of short- and long-term trends is more theoretically robust.
- Linear trend statistics, though important for understanding long-term change processes, do not offer a direct defence for the core theory of global warming. They are a tool.
- Theoretical understanding is separated from methodological issues affecting detection, attribution, forecasting and risk, allowing each to be dealt with in a fit-for-purpose manner (Jones, 2015a).

A significant unknown is what happens to the additional heat energy due to anthropogenic greenhouse gases between the time it is trapped and is measured as warming (Figure 2). This is a known unknown that sits between two well-understood aspects of core theory. How much of that warming is diffuse (gradual), how much is entrained into quasi-periodic oscillations and what processes govern the delay? This paper puts forward a partial explanation for what the process may be. Although openly admitting to this ignorance would invite an immediate response from climate's critics about "what else doesn't climate science know?", the core theory is extremely robust, and therefore a clear defence exists.

Conclusions

The title of the paper asks whether both sides arguing the so-called pause, are wrong. The paper shows that in some ways they are, and either side is correct in other ways, though not equally so. As documented here and in Jones and Ricketts (2016), step-like changes are being detected in a wide range of climate and climate-related variables. Their attribution to random climate variability on the one hand and to non-greenhouse gas external drivers on the other both support *H1*, the first explicitly and the second implicitly. There is a small body of scientific opinion supporting the construct in Figure 1c, where interactions between change and variability produce a nonlinear signal on decadal timescales, integrating into a long-term complex trend. This supports *H2*, as do the conclusions of Jones and Ricketts (2016) who apply the severe testing methodology outlined above.

If *H2* is true, Skeptical Science's Realists are correct in their defence of the underpinning science but incorrect in the characterisation of the underlying signal as gradual. This underdetermines risk. They are also incorrect in their rejection of step changes as being noise and not contributing to the signal. The Contrarians may have detected some step-like changes in warming data but are incorrect in their assertion that it challenges the theory in any way, or leads to a downgrading of future risk. If *H2* is true, most of the existing climate theory remains intact but has the potential to become better integrated by reconciling the linear and nonlinear responses to change. However, methods for detection and attribution, the characterisation of risk and what is communicated would need to change, because they are strongly based on *H1*. Scientific narratives would also need to change from the existing gradualist narrative to one that addresses complexity and communicates uncertainty more effectively.

How the climate warms on decadal timescales is too important an issue to be caught up in the climate wars and fought over using simple statistical inferences. Here, we have tried to provide a philosophically-sound defence for climate science, while exploring how to reconcile externally-forced change and internally-generated variability using severe testing.

References

Aires, F., C. Prigent, W. Rossow and M. Rothstein, 2001: A new neural network approach including first-guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations. *Journal of Geophysical Research: Atmospheres*, **106**, 14887-14907.

Arthur, W.B., 2015: *Complexity and the Economy*. Oxford University Press, New York, 211 pp.

Bartsev, S.I., P.V. Belolipetskii, A.G. Degermendzhi, Y.D. Ivanova, A.A. Pochekutov and M.Y. Saltykov, 2016: Refocusing on the dynamics of the Earth's climate. *Herald of the Russian Academy of Sciences*, **86**, 135-142.

Beaulieu, C., J. Chen and J.L. Sarmiento, 2012: Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **370**, 1228-1249.

Belolipetsky, P.V., 2014: The Shifts Hypothesis - an alternative view of global climate change. *arXiv preprint arXiv:1406.5805*.

Beltrami, H., S.M. Gurpreet and E.S. Jason, 2015: Ground surface temperature and continental heat gain: uncertainties from underground. *Environmental Research Letters*, **10**, 014009.

Benestad, R.E., 2016: A mental picture of the greenhouse effect. *Theoretical and Applied Climatology*, 1-10.

Branstator, G. and F. Selten, 2009: "Modes of Variability" and Climate Change. *Journal of Climate*, **22**, 2639-2658.

Cahill, N., S. Rahmstorf and A.C. Parnell, 2015: Change points of global temperature. *Environmental Research Letters*, **10**, 084002.

Carter, B., 2006: *There IS a problem with global warming... it stopped in 1998*. Telegraph Media Group Limited, London, UKpp.

Carter, R.M., C. De Freitas, I.M. Goklany, D. Holland, R.S. Lindzen, I. Byatt, I. Castles, D. Henderson, N. Lawson and R. McKitrick, 2006: The Stern review: a dual critique. *World Economics*, **7**, 165-232.

Chikamoto, Y., M. Kimoto, M. Watanabe, M. Ishii and T. Mochizuki, 2012: Relationship between the Pacific and Atlantic stepwise climate change during the 1990s. *Geophysical Research Letters*, **39**.

Coady, D. and R. Corry, 2013: *The Climate Change Debate: An Epistemic and Ethical Enquiry*. Palgrave Macmillan UK, London, 122 pp.

Cohn, T.A. and H.F. Lins, 2005: Nature's style: Naturally trendy. *Geophysical Research Letters*, **32**, L23402.

Collins, M., T. Fricker and L. Hermanson, 2011: From observations to forecasts – Part 9: what is decadal forecasting? *Weather*, **66**, 160-164.

Compo, G.P. and P.D. Sardeshmukh, 2009: Oceanic influences on recent continental warming. *Climate Dynamics*, **32**, 333-342.

Corti, S., F. Molteni and T.N. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799-802.

Cox, D. and D.G. Mayo, 2010: Objectivity and Conditionality in Frequentist Inference. In: *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* [Mayo, D.G. and A. Spanos (eds.)] Cambridge University Press, Cambridge UK and New York USA, 276-304.

Curd, M. and J.A. Cover, 1998: *Philosophy of Science: The Central Issues*. WW Norton New York and London, 1379 pp.

Deser, C., A. Phillips, V. Bourdette and H. Teng, 2012: Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, **38**, 527-546.

Dommenget, D., 2009: The Ocean's Role in Continental Climate Variability and Change. *Journal of Climate*, **22**, 4939-4952.

Douglass, D. and J. Christy, 2009: Limits on CO2 climate forcing from recent temperature data of Earth. *Energy & Environment*, **20**, 177-189.

Douglass, D. and R. Knox, 2012: Ocean heat content and Earth's radiation imbalance. II. Relation to climate shifts. *Physics Letters A*, **376**, 1226-1229.

Douglass, D.H. and R.S. Knox, 2015a: The Sun is the climate pacemaker II. Global ocean temperatures. *Physics Letters A*, **379**, 830-834.

Douglass, D.H. and R.S. Knox, 2015b: The Sun is the climate pacemaker I. Equatorial Pacific Ocean temperatures. *Physics Letters A*, **379**, 823-829.

Ebi, K.L., 2011: Differentiating theory from evidence in determining confidence in an assessment finding. *Climatic Change*, **108**, 693-700.

Edwards, P.N., 1999: Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture*, **8**, 437 - 472.

Enfield, D.B. and L. Cid-Serrano, 2006: Projecting the risk of future climate shifts. *International Journal of Climatology*, **26**, 885-895.

Fasullo, J.T. and K.E. Trenberth, 2008a: The Annual Cycle of the Energy Budget. Part I: Global Mean and Land–Ocean Exchanges. *Journal of Climate*, **21**, 2297-2312.

Fasullo, J.T. and K.E. Trenberth, 2008b: The Annual Cycle of the Energy Budget. Part II: Meridional Structures and Poleward Transports. *Journal of Climate*, **21**, 2313-2325.

Fischer, T., M. Gemmer, L. Liu and B. Su, 2012: Change-points in climate extremes in the Zhujiang River Basin, South China, 1961–2007. *Climatic Change*, **110**, 783-799.

Geoffroy, O., D. Saint-Martin and A. Voldoire, 2015: Land-sea warming contrast: the role of the horizontal energy transport. *Climate Dynamics*, **45**, 3493-3511.

Gervais, F., 2016: Anthropogenic CO2 warming challenged by 60-year cycle. *Earth-Science Reviews*, **155**, 129-135.

Gigerenzer, G. and J.N. Marewski, 2015: Surrogate Science: The Idol of a Universal Method for Scientific Inference. *Journal of Management*, **41**, 421-440.

Hare, S.R. and N.J. Mantua, 2000: Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress In Oceanography*, **47**, 103-145.

Hasselmann, K., 1976: Stochastic climate models part I. Theory. *Tellus*, **28**, 473-485.

Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. In: *Meteorology of Tropical Oceans* [Shaw, D.B. (ed.) Royal Meteorological Society, London, UK, 251-259.

Hasselmann, K., 2002: Is Climate Predictable? In: *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes* [Bunde, A., J. Kropp and H.J. Schellnhuber (eds.)] Springer, Berlin Heidelberg, 141-188.

Hegerl, G.C., F.W. Zwiers, P. Braconnot, N.P. Gillett, Y. Luo, J.A.M. Orsini, N. Nicholls, J.E. Penner and P.A. Stott, 2007: Understanding and Attributing Climate Change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)] Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 663-745.

Held, I.M., 2005: The Gap between Simulation and Understanding in Climate Modeling. *Bulletin of the American Meteorological Society*, **86**, 1609-1614.

Hope, P., B. Timbal and R. Fawcett, 2010: Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time. *International Journal of Climatology*, **30**, 1360-1372.

IPCC, 1990: *Climate Change: the IPCC Scientific Assessment*. Cambridge University Press, Cambridge UK, New York USA, 365 pp.

IPCC, 1996: *Climate Change 1995 : the Science of Climate Change*. Cambridge University Press, Cambridge ; New York, 572 pp.

IPCC, 2001a: *Climate Change 2001: Synthesis Report*. Cambridge University Press, Cambridge, U.K., 397 pp.

IPCC, 2001b: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.

IPCC, 2007: *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 996 pp.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

IPCC, 2014a: *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva Switzerland, 151 pp.

IPCC, 2014b: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USApp.

Jones, R.N., 2012: Detecting and attributing nonlinear anthropogenic regional warming in southeastern Australia. *Journal of Geophysical Research*, **117**, D04105.

Jones, R.N., C.K. Young, J. Handmer, A. Keating, G.D. Mekala and P. Sheehan, 2013: *Valuing Adaptation under Rapid Change*. National Climate Change Adaptation Research Facility, Gold Coast, Australia, 182 pp.

Jones, R.N., 2015a: *Reconciling anthropogenic climate change and variability on decadal timescales: the challenge*. Climate Change Working Paper No. 36, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 15 pp.

Jones, R.N., 2015b: *Reconciling anthropogenic climate change and variability on decadal timescales: history and philosophy*. Climate Change Working Paper No. 33, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 18 pp.

Jones, R.N. and J.H. Ricketts, 2015: *Analysing steps in global and regional observed air temperature*. Climate Change Working Paper No. 34, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 20 pp.

Jones, R.N. and J.H. Ricketts, 2016: *Reconciling the signal and noise of atmospheric warming on decadal timescales*. Climate Change Working Paper No. 38, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 65 pp.

Katzav, J., 2013: Severe testing of climate change hypotheses. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **44**, 433-441.

Kirtman, B., S. Power, A.J. Adedoyin, G. Boer, R. Bojariu, I. Camilloni, F. Doblas-Reyes, A. Fiore, M. Kimoto, G. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J.v. Oldenborgh, G. Vecchi and H.-J. Wang, 2013: Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013: The Physical Science Basis. Working Group I contribution to the IPCC 5th Assessment Report* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)] Cambridge University Press, Cambridge and New York, 121.

Koutsoyiannis, D., 2010: HESS Opinions "A random walk on water". *Hydrology and Earth System Sciences*, **14**, 585-601.

Kuhn, T.S., 1970: Reflections on my critics. In: *Criticism and the Growth of Knowledge* [Lakatos, I. and A. Musgrave (eds.)] Cambridge University Press, Cambridge UK, 231-278.

Kuhn, T.S., 1977: Objectivity, value judgment, and theory choice. In: *The Essential Tension: Selected Studies in Scientific Tradition and Change* University of Chicago Press, Chicago, 320-39.

Kuhn, T.S., 1996: *The Structure of Scientific Revolutions*, 3rd Edn. University of Chicago Press, Chicago, IL, 212 pp.

Lakatos, I., 1970: Falsification and the methodology of science research programmes. In: *Criticism and the growth of knowledge* [Lakatos, I. and A. Musgrave (eds.)] University Press, Cambridge [Eng.], 91-196.

Lakatos, I., 1971: History of Science and its Rational Reconstructions. In: *PSA 1970: In Memory of Rudolf Carnap Proceedings of the 1970 Biennial Meeting Philosophy of Science Association* [Buck, R.C. and R.S. Cohen (eds.)] Springer Netherlands, Dordrecht, 91-136.

Lakatos, I., 1978: *The Methodology of Scientific Research Programmes*. Cambridge University Press, Cambridge ; New York, 250 pp.

Lambert, F.H., M.J. Webb and M.M. Joshi, 2011: The Relationship between Land–Ocean Surface Temperature Contrast and Radiative Forcing. *Journal of Climate*, **24**, 3239-3256.

Laudan, L., 1977: *Progress and its Problems: Toward a Theory of Scientific Growth*. University of California Press, Berkeley, 257 pp.

Legates, D.R., W. Soon and W.M. Briggs, 2015: Climate consensus and 'misinformation': A rejoinder to Agnotology, scientific consensus, and the teaching and learning of climate change. *Science & Education*, **24**, 299-318.

Levitus, S., J.I. Antonov, T.P. Boyer, O.K. Baranova, H.E. Garcia, R.A. Locarnini, A.V. Mishonov, J.R. Reagan, D. Seidov, E.S. Yarosh and M.M. Zweng, 2012: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, **39**, L10603.

Lewandowsky, S., J.S. Risbey and N. Oreskes, 2016: The "Pause" in Global Warming: Turning a Routine Fluctuation into a Problem for Science. *Bulletin of the American Meteorological Society*, **97**, 723–733.

Lloyd, E.A., 2012: The role of 'complex'empiricism in the debates about satellite data and climate models. *Studies in History and Philosophy of Science Part A*, **43**, 390-401.

Longino, H.E., 1990: *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press, Princeton, NJ, 262 pp.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20**, 130-141.

Lorenz, E.N., 1975: Climate Predictability. In: *The Physical Bases of Climate and Climate Modelling.*, Vol. GARP Publication Series, Vol. 16 World Meteorological Organisation, Geneva, 132-136.

Lucarini, V., K. Fraedrich and F. Ragone, 2010: New results on the thermodynamical properties of the climate system. *arXiv preprint arXiv:1002.0157*.

Lucarini, V. and F. Ragone, 2011: Energetics of climate models: net energy balance and meridional enthalpy transport. *Reviews of Geophysics*, **49**, RG1001.

Mantua, N. and S. Hare, 2002: The Pacific Decadal Oscillation. *Journal of Oceanography*, **58**, 35-44.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace and R.C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, **78**, 1069-1079.

Mayo, D.G., 1996: *Error and the Growth of Experimental Knowledge*. University of Chicago Press, Chicago, 509 pp.

Mayo, D.G., 2005: Evidence as passing severe tests: highly probable versus highly probed hypotheses. In: *Scientific Evidence: Philosophical Theories and Applications* [Achinstein, P. (ed.) John Hopkins University Press, Baltimore and London, 95-127.

Mayo, D.G. and D. Cox, 2010: Frequentist Statistics as a Theory of Inductive Inference. In: *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* [Mayo, D.G. and A. Spanos (eds.)] Cambridge University Press, Cambridge UK and New York USA, 247-275.

Mayo, D.G. and A. Spanos, 2010: *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of science*. Cambridge University Press, Cambridge UK, 419 pp.

Mayo, D.G. and A. Spanos, 2011: Error Statistics. In: *Philosophy of Statistics* [Bandyopadhyay, P.S. and M.R. Forster (eds.)], Vol. 7 North-Holland, Amsterdam, 153-198.

McKitrick, R.R., 2014: HAC-Robust Measurement of the Duration of a Trendless Subsample in a Global Climate Time Series. *Open Journal of Statistics*, **4**, 527-535.

McKitrick, R.R., 2015: Energy Policy and Environmental Stewardship: Risk Management not Risk Avoidance. *Greer-Heard Point-Counterpoint Forum New Orleans Baptist Theological Seminary, April 10-11 2015*, City, pp 18.

Meehl, G.A., A. Hu and B.D. Santer, 2009: The Mid-1970s Climate Shift in the Pacific and the Relative Roles of Forced versus Inherent Decadal Variability. *Journal of Climate*, **22**, 780-792.

Meehl, G.A., L. Goddard, G. Boer, R. Burgman, G. Branstator, C. Cassou, S. Corti, G. Danabasoglu, F. Doblas-Reyes, E. Hawkins, A. Karspeck, M. Kimoto, A. Kumar, D. Matei, J. Mignot, R. Msadek, A. Navarra, H. Pohlmann, M. Rienecker, T. Rosati, E. Schneider, D. Smith, R. Sutton, H. Teng, G.J. van Oldenborgh, G. Vecchi and S. Yeager, 2013: Decadal Climate Prediction: An Update from the Trenches. *Bulletin of the American Meteorological Society*, **95**, 243-267.

Menne, M., 2006: Abrupt global temperature change and the instrumental record. *18th Conference on Climate Variability and Change, Atlanta, Georgia*, American Meteorological Organisation, City, pp.

Mercer, D., 2016: Why Popper can't resolve the debate over global warming: Problems with the uses of philosophy of science in the media and public framing of the science of global warming. *Public Understanding of Science*.

Miller, D., 1974: On the comparison of false theories by their bases. *British Journal for the Philosophy of Science*, 178-188.

Musgrave, A., 2010: Critical Rationalism, Explanation and Severe Tests. In: *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* [Mayo, D.G. and A. Spanos (eds.)] Cambridge University Press, Cambridge UK, 88-112.

Newman, M., G.P. Compo and M.A. Alexander, 2003: ENSO-forced variability of the Pacific decadal oscillation. *Journal of Climate*, **16**, 3853-3857.

Nicholls, N., P. Dellamarta and D. Collins, 2004: 20th century changes in temperature and rainfall in New South Wales. *Australian Meteorological Magazine*, **53**, 263-268.

O'Kane, T.J., R.J. Matear, M.A. Chamberlain and P.R. Oke, 2014: ENSO regimes and the late 1970's climate shift: The role of synoptic weather and South Pacific ocean spiciness. *Journal of Computational Physics*, **271**, 19-38.

Overland, J., S. Rodionov, S. Minobe and N. Bond, 2008: North Pacific regime shifts: Definitions, issues and recent transitions. *Progress In Oceanography*, **77**, 92-102.

Ozawa, H., A. Ohmura, R.D. Lorenz and T. Pujol, 2003: The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle. *Reviews of Geophysics*, **41**, 1018.

Palmer, T., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, **74**, 49-65.

Palmer, T., 1998: Nonlinear dynamics and climate change: Rossby's legacy. *Bulletin of the American Meteorological Society*, **79**, 1411-1423.

Palmer, T.N., 1999: A nonlinear dynamical perspective on climate prediction. *Journal of Climate*, **12**, 575-591.

Pierrehumbert, R.T., 2011: Infrared radiation and planetary temperature. *Physics Today*, **64**, 33-38.

Plimer, I., 2009: *Heaven and Earth: Global Warming: the Missing Science*. Connor Court Publishing, Ballan, Australia, 504 pp.

Popper, K.R., 1992: *The Logic of Scientific Discovery*. Routledge, London; New York, 479 pp.

Popper, K.R., 2002: *Conjectures and Refutations: the Growth of Scientific Knowledge*. Routledge, London; New York, 582 pp.

Raper, S.C.B., J.M. Gregory and T.J. Osborn, 2001: Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Climate Dynamics*, **17**, 601-613.

Raper, S.C.B., J.M. Gregory and R.J. Stouffer, 2002: The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *Journal of Climate*, **15**, 124-130.

Reeves, J., J. Chen, X.L. Wang, R. Lund and Q.Q. Lu, 2007: A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, **46**, 900-915.

Reid, P.C. and G. Beaugrand, 2012: Global synchrony of an accelerating rise in sea surface temperature. *Journal of the Marine Biological Association of the United Kingdom*, **92**, 1435-1450.

Ricketts, J.H., 2015: A probabilistic approach to climate regime shift detection based on Maronna's bivariate test. *The 21st International Congress on Modelling and Simulation (MODSIM2015), Gold Coast, Queensland, Australia*, Weber, T., M.J. McPhee and R.S. Anderssen, Eds., The Modelling and Simulation Society of Australia and New Zealand, City, pp 1310-1316.

Ricketts, J.H. and R.N. Jones, 2016: The multi-step Maronna-Yohai bivariate test for detecting multiple step changes in climate data. *Manuscript in preparation*.

Risbey, J., 2015: Earth science: Free and forced climate variations. *Nature*, **517**, 562-563.

Rodionov, S.N., 2005: A brief overview of the regime shift detection methods. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability. UNESCO-ROSTE/BAS Workshop on Regime Shifts, Varna, Bulgaria*, Velikova, V. and N. Chipev, Eds., City, pp 17-24.

Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton and S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. *Nature Climate Change*, **5**, 240-245.

Roush, S., 2005: *Tracking truth: knowledge, evidence, and science*. Clarendon Press, Oxford, 235 pp.

Ruggieri, E., 2013: A Bayesian approach to detecting change points in climatic records. *International Journal of Climatology*, **33**, 520-528.

Santer, B. and C. Mears, 2015: *Human Effects on Climate are Reality, not Science Fiction*, The Governor's Office of Planning and Research, California, 4 pp.

Santer, B.D., C. Mears, C. Doutriaux, P. Caldwell, P.J. Gleckler, T.M.L. Wigley, S. Solomon, N.P. Gillett, D. Ivanova, T.R. Karl, J.R. Lanzante, G.A. Meehl, P.A. Stott, K.E. Taylor, P.W. Thorne, M.F. Wehner and F.J. Wentz, 2011: Separating signal and noise in atmospheric temperature changes: The importance of timescale. *Journal of Geophysical Research*, **116**, D22105.

Seidel, D.J. and J.R. Lanzante, 2004: An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *Journal of Geophysical Research*, **109**, D14108.

Semenov, V.A., M. Latif, D. Dommenget, N.S. Keenlyside, A. Strehz, T. Martin and W. Park, 2010: The Impact of North Atlantic–Arctic Multidecadal Variability on Northern Hemisphere Surface Air Temperature. *Journal of Climate*, **23**, 5668-5677.

Şerban, D.Z. and B.H. Jacobsen, 2001: The use of broad-band prior covariance for inverse palaeoclimate estimation. *Geophysical Journal International*, **147**, 29-40.

Singer, S.F., 2011: Lack of Consistency between Modeled and Observed Temperature Trends. *Energy & Environment*, **22**, 375-406.

Skeptical Science, 2015: *The Escalator*. John Cook, Brisbane, http://www.skepticalscience.com/graphics.php?g=47, (Last accessed February 23, 2016).

Slingo, J., 2013: *Statistical models and the global temperature record*, UK Met Office, Exeter, Devon, 15 pp.

Sloan, T. and A.W. Wolfendale, 2013: Cosmic rays, solar activity and the climate. *Environmental Research Letters*, **8**, 045022.

Solomon, A., L. Goddard, A. Kumar, J. Carton, C. Deser, I. Fukumori, A.M. Greene, G. Hegerl, B. Kirtman, Y. Kushnir, M. Newman, D. Smith, D. Vimont, T. Delworth, G.A. Meehl and T. Stockdale, 2011: Distinguishing the Roles of Natural and Anthropogenically Forced Decadal Climate Variability. *Bulletin of the American Meteorological Society*, **92**, 141-156.

Spanos, A., 2010a: Theory Testing in Economics and the Error Statistical Perspective. In: *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* [Mayo, D.G. and A. Spanos (eds.)] Cambridge University Press, Cambridge UK, 202-246.

Spanos, A., 2010b: Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics*, **158**, 204-220.

Stevens, B. and S.E. Schwartz, 2012: Observing and Modeling Earth's Energy Flows. *Surveys in Geophysics*, **33**, 779-816.

Symmons, P., 2011: Temperature Estimates from Models and Observations. *Energy & Environment*, **22**, 1059-1068.

Thornton, S., 2015: *Karl Popper*, Winter 2015 Edn. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University Stanfordpp.

Tichý, P., 1974: On Popper's definitions of verisimilitude. *British Journal for the Philosophy of Science*, 155-160.

Tisdale, B., 2015: *On Global Warming and the Illusion of Control – Part 1. A Comprehensive Illustrated Introduction to the Hypothesis of Human-Induced Global Warming*. Bob Tisdale, bobtisdale.wordpress.com, 733 pp.

Trenberth, K.E., 2011: Changes in precipitation with climate change. *Climate Research*, **47**, 123.

Trenberth, K.E., 2015: Has there been a hiatus? *Science*, **349**, 691-692.

Tsonis, A. and K. Swanson, 2012: Review article" On the origins of decadal climate variability: a network perspective". *Nonlinear Processes in Geophysics*, **19**, 559-568.

Tsonis, A.A. and K.L. Swanson, 2011: Climate mode covariability and climate shifts. *International Journal of Bifurcation and Chaos*, **21**, 3549-3556.

van Fraassen, B., 2008: *Scientific Representation: Paradoxes of Perspective*. Clarendon Press, Oxford, 303 pp.

Vivès, B. and R.N. Jones, 2005: *Detection of Abrupt Changes in Australian Decadal Rainfall (1890-1989)*. CSIRO Atmospheric Research Technical Paper, CSIRO Atmospheric Research, Melbourne, 54 pp.

von Schuckmann, K., M.D. Palmer, K.E. Trenberth, A. Cazenave, D. Chambers, N. Champollion, J. Hansen, S.A. Josey, N. Loeb, P.P. Mathieu, B. Meyssignac and M. Wild, 2016: An imperative to monitor Earth's energy imbalance. *Nature Clim. Change*, **6**, 138-144.

von Storch, H., 1999: Misuses of Statistical Analysis in Climate Research. In: *Analysis of Climate Variability* [von Storch, H. and A. Navarra (eds.)] Springer Berlin Heidelberg, 11-26.

von Storch, H. and F.W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge UK, 484 pp.

Warner, R.F., 1995: Predicting and managing channel change in Southeast Australia. *Catena*, **25**, 403-418.

Weart, S.R., 2008: *The Discovery of Global Warming*. Harvard University Press, Cambridge, MA, 240 pp.

Wigley, T.M.L., 2008: *MAGICC/SCENGEN 5.3: User Manual (Version 2)*, National Center for Atmosperic Research (NCAR), Boulder, Colorado, 80 pp.

Winton, M., K. Takahashi and I.M. Held, 2010: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change. *Journal of Climate*, **23**, 2333-2344.

Zhang, R. and T.L. Delworth, 2007: Impact of the Atlantic multidecadal oscillation on North Pacific climate variability. *Geophysical Research Letters*, **34**.

Zimmerman, M.E., 2009: Including and differentiating among perspectives: An integral approach to climate change. *Journal of Integral Theory & Practice*, **4**, 1-26.