



24 Critical Errors in Tol (2014)

Reaffirming the 97% consensus
on anthropogenic global warming

Authors

John Cook^{1,2,3}, Dana Nuccitelli², Andy Skuce⁴, Robert Way⁵, Peter Jacobs⁶, Rob Painting², Rob Honeycutt², Sarah A. Green⁷, Stephan Lewandowsky^{8,3}, Alexander Coulter⁹

Affiliations

¹ Global Change Institute, The University of Queensland, Australia

² Skeptical Science, Brisbane, Queensland, Australia

³ School of Psychology, University of Western Australia, Australia

⁴ Salt Spring Consulting Ltd, Salt Spring Island, BC, Canada

⁵ Department of Geography, University of Ottawa, Canada

⁶ Department of Environmental Science and Policy, George Mason University, Fairfax, VA, USA

⁷ Department of Chemistry, Michigan Technological University, Houghton, Michigan

⁸ School of Experimental Psychology and Cabot Institute, University of Bristol, UK

⁹ Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, MI, USA

Corresponding author:

John Cook (j.cook3@uq.edu.au)

Acknowledgements

Thanks to the following for their comments during the writing of this document: Collin JHM Maessen, Bärbel Winkler, George Morrison.

First published in June 2014.

For more information or to comment on this report, visit <http://sks.to/tolreply>

24 Critical Errors in Tol (2014): Reaffirming the 97% consensus on anthropogenic global warming is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License. Extracts may be reproduced provided Skeptical Science is attributed with a link to www.skepticalscience.com.

“

“There is no doubt in my mind that the literature on climate change overwhelmingly supports the hypothesis that climate change is caused by humans. I have very little reason to doubt that the consensus is indeed correct.”

”

Richard Tol

The strengthening scientific consensus on global warming

A number of studies have found evidence for a scientific consensus on anthropogenic global warming (AGW). Oreskes (2004) analysed 928 'global climate change' papers published from 1993 to 2003 and found none rejecting AGW. A survey of Earth scientists found that among climate scientists who actively publish climate research, 97% agreed that humans are significantly increasing global temperature (Doran and Zimmermann, 2009). An analysis of public statements by publishing climate scientists also found 97% agreement with the scientific consensus on AGW (Anderegg et al., 2011).

Three independent studies (Doran & Zimmermann, 2009; Anderegg et al. 2011, Cook et al. 2013), using three different methods, have found a 97% scientific consensus that humans are causing global warming.



Repeated surveys of scientists found that scientific agreement about AGW steadily increased from 1996 to 2009 (Bray 2010). The consensus is reflected in the increasingly definitive statements issued by the Intergovernmental Panel on Climate Change on the attribution of recent global warming (Houghton et al 1996, 2001, Solomon et al 2007; IPCC 2013, 2014). Citation analysis has also found that consensus on AGW formed and strengthened from the early 1990s (Shwed and Bearman, 2010).

Cook et al., 2013 (C13) conducted an analysis of 'global climate change' and 'global warming' papers published from 1991 to 2011. C13 found that among abstracts stating a position on AGW, 97.1% endorsed AGW. The authors of the papers were also invited to self-rate their own papers. Among papers self-rated as stating a position on AGW, 97.2% endorsed AGW.

This report documents 25 errors in T14 that in sum falsify its conclusions.

Tol, 2014 (T14) agrees that "the literature on climate change overwhelmingly supports the hypothesis that climate change is caused by humans" but disputes the methods used in C13. However, T14 contains a number of critical errors that falsify key conclusions, include fundamental mathematical mistakes, use inappropriate statistics, and make unsubstantiated assertions (Cook et al., 2014 or C14).

This report documents 24 errors in T14 that falsify its conclusions.

**Error
1**

Flawed attempt to recalculate consensus

“ T14 Sec. 3.3.1: “Reconciliations and reratings were biased towards a rejection of the hypothesis of anthropogenic climate change ($\chi^2=62$; $p<0.001$; Figure S20). However, the number of endorsements far exceeds the number of rejections. Therefore, applying the same correction to the 6.7% incorrectly rated abstracts, the consensus rate falls from 98% to 91%.”

T14 Sec 3.4: “If methods and palaeoclimate papers are misrated in the same proportion as impacts and mitigation papers, then the consensus rate is 89.9% (all endorsements) and 93.8% (explicit endorsements only).”

T14 Sec. 4.0: “Removing irrelevant papers, I find that, rather than 3%, up to 10% of papers explicitly disagree with the hypothesis that climate change is real and largely anthropogenic.”

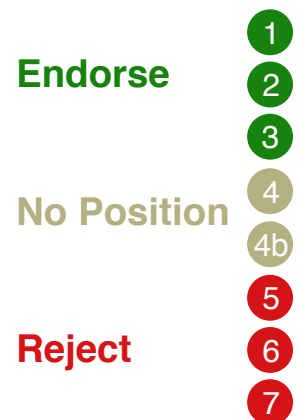
C13 classified abstracts of climate science papers based on the level of endorsement that most of the recent global warming is man-made (AGW, Categories 1–3), rejection or minimisation of AGW (Categories 5–7), or ‘no position’ on AGW (Category 4). Abstracts taking a position on AGW (Categories 1–3 & 5–7, plus an estimated 40 ‘undecided’ papers) were used to calculate the consensus; 97.1% endorsed AGW. Each abstract was categorised by at least two independent raters, and a reconciliation process resolved rater disagreements.

Healey (2011) highlights the problematic approach of using an ordinal variable (e.g., level of endorsement labelled 1 to 7) to make inferences about a continuous quantity such as consensus (in this case, defined as the percentage of endorsements among papers stating a position on AGW). We demonstrate that statistics based on inappropriate use of ordinal measures cannot be used to infer uncertainties in the consensus.

T14 applies a ‘correction’ assuming that the probability of an erroneous rating and the size of correction does not depend on the rating category, except for preventing changes to outside of the category range 1–7. This assumption is incorrect and invalidates the T14 correction. Disagreement was much more common over endorsement or rejection ratings (34% of disputed cases) than over no position ratings (9% of cases). T14 assumes, based on the average of all reconciliations, that among ‘no position’ papers that are misrated, 55% become ‘rejection’ papers and 45% become ‘endorsement’ papers. In reality, during reconciliation, 2% of ‘no position’ abstracts were moved to ‘rejection’ and 98% to ‘endorsement’ (Figure 1a).

The false assumption of applying the same correction for all categories leads T14 to recategorise 293 “No Position” abstracts as Rejection abstracts. This more than quadruples the number of rejection abstracts from 78 to 379. However, T14 does not identify a single one of the supposed 300 extra rejection abstracts. This is the primary contributor to Tol’s unjustified and consequently incorrect adjustment of the 97% consensus to 91% (Figure 1b).

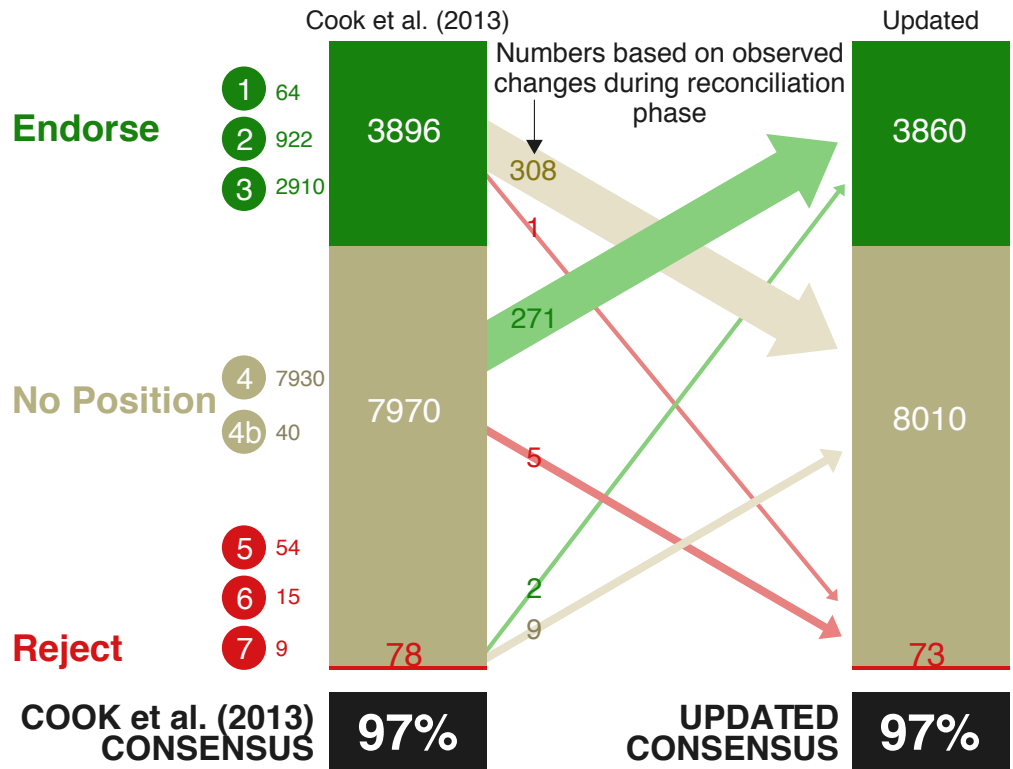
Endorsement Levels



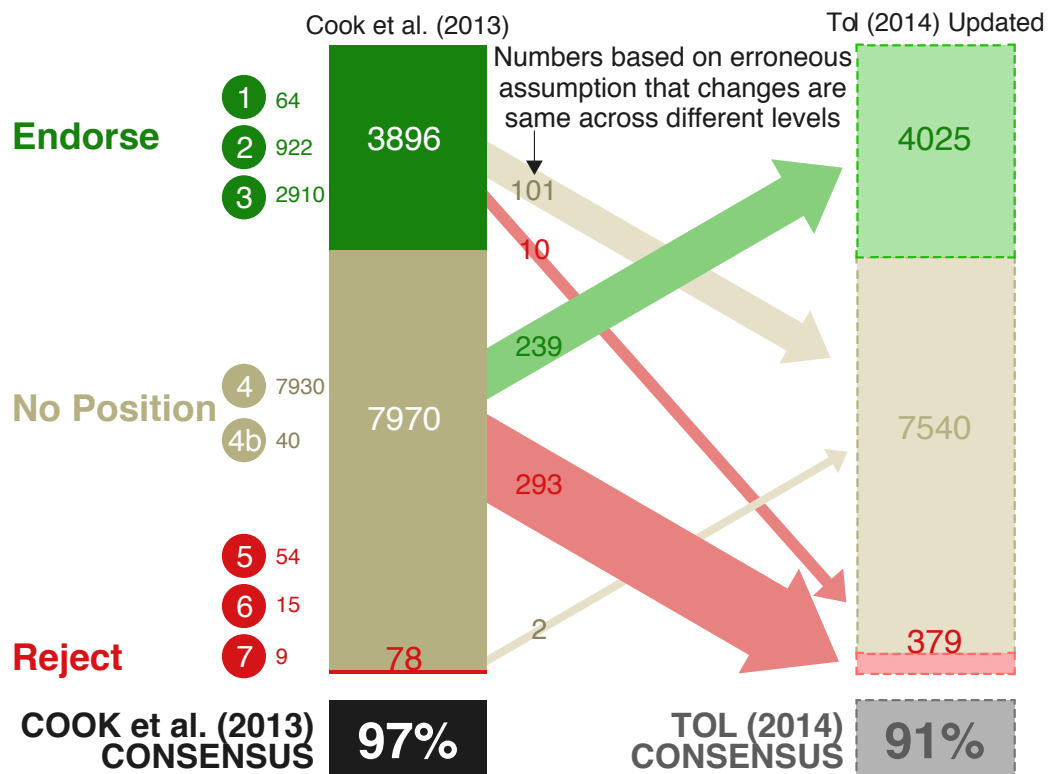
T14 does not identify a single one of the supposed 300 extra rejection abstracts.

Figure 1. Changes in initial to final abstract ratings assuming a 6.7% uncertainty. a) Recalculated consensus value based on actual proportional endorsement changes during the resolution process. b) Tol's method of recalculating consensus, based on the erroneous assumption that all endorsement levels change at the same rate. This assumption changes 293 "No Position" abstracts to "Rejection" abstracts.

a) Recalculation of consensus based on observed changes (different endorsement levels change at different rates)



b) Tol (2014) erroneously assumes all endorsement levels change in the same way, in contradiction to observations



Using the observed changes during reconciliation to perform the correction results in the consensus changing from 97.1% to 97.2%. T14's claim of a reduction to 91% is consequently the result of a basic calculation error.

Changes within a grouped category (e.g. from within endorsements {1,2,3} to another category within {1,2,3}) do not affect the consensus. Figure 1 in C14 (reproduced above) shows the number that would be expected to change between or within each of the category groups representing 'endorsement', 'no position' or 'rejection'. Using the observed changes during reconciliation to perform the correction results in the consensus changing from 97.1% to 97.2%. T14's claim of a reduction to 91% is consequently the result of a basic calculation error.

Error

2

Mischaracterises key result of C13

“ T14 Abstract: “A claim has been that 97% of the scientific literature endorse anthropogenic climate change”.

C13 made no claims about the entirety of the scientific literature. Its key result is that 97% of the papers on ‘global climate change’ or ‘global warming’ that state a position on anthropogenic global warming endorsed the consensus position. This clarification is important because C13’s results have previously been mischaracterised in this fashion, failing to distinguish between all climate papers (or even all scientific papers) and papers stating a position on AGW.

One significance of the distinction between papers that do or do not state a position on AGW is discussed in C13 (Section 4). Oreskes (2007) predicted that as the consensus strengthens, one expects to see a smaller proportion of papers explicitly endorsing AGW as research moves on to address other questions that remain unsettled. Such a temporal evolution of scholarly research has been characterised as following a “spiral trajectory”, and indeed is expected behavior (Shwed and Bearman, 2010). This prediction was observed in C13.

This is a mischaracterisation of a key result of C13, which found that 97% of ‘global climate change’ or ‘global warming’ papers’ stating a position on anthropogenic global warming endorsed the consensus position.

Error

3

Neglects scholarly literature on consensus

“ T14 Intro: “consensus has no academic value (although the occasional stock take is valuable for teaching and guiding future research) and limited policy value.”

This claim is wholly unsubstantiated and is refuted by decades of academic research on scientific consensus from historical, sociological, philosophical, and other perspectives (Knorr, 1978; Mulkay, 1978; Lehrer and Wagner, 1981; Kim, 1994; Shwed and Bearman, 2010; Miller, 2013).

Although C13 introduced novel information about the scientific consensus on AGW (surveying an unprecedented number of abstracts, detailing its temporal evolution, soliciting self-ratings from authors, etc.), other studies have found similar widespread expert agreement, among them Oreskes (2004). While the value of citation count as a tool to assess research(er) impact and quality is a hotly debated subject, citation count has been shown to be correlated with the importance of a paper (Abt, 2000; Buéla-Casal and Zych, 2010), and Tol himself places a great deal of import on citation numbers (Tol, 2014a.; Tol, n.d.). By this metric, the scholarly consensus on anthropogenic warming is strikingly important, with Oreskes (2004) receiving a remarkable number of citations (250 to date, according to Thomson

This claim is wholly unsubstantiated and is self-evidently refuted by decades of academic research on scientific consensus from historical, sociological, philosophical, and other perspectives.

Consensus is of particular interest and importance at the science-policy interface, where it plays a crucial role in decision-making with respect to management issues that involve a scientific component.

Reuters' Web of Science™; for perspective, Tol's most-cited paper as lead author has received 235 citations to date in the same index).

Further, consensus is of particular interest and importance at the science-policy interface, where it plays a crucial role in decision-making with respect to management issues that involve a scientific component (e.g. Boesch, 1999; NCEAS, 2001; Heisler et al., 2008).

A growing body of research has demonstrated that for climate change the scientific consensus is relevant to policy. Public understanding of the extent of scientific consensus on climate change is fundamental to acceptance of its threats and anthropogenic cause (Ding et al., 2011; Lewandowsky et al., 2012; Kotcher et al., 2014). Moreover, perceived scientific consensus is one of the strongest predictors of support for climate policy (Ding et al., 2011; McCright et al., 2013). Conversely, perceived scientific disagreement can severely diminish public belief that environmental problems are occurring and that they require policy response (Aklin and Urpelainen, 2014).

Finally, it should be noted that attempted consensus devaluation is a staple of contrarian and science-denialist rhetoric. Creationists and "Intelligent Design" advocates are notorious for claiming that consensus is of little value in science or is itself anti-scientific (Behe, 2009; Meyer, 2009; C4ID, 2011). Groups pushing discredited, contrarian positions with respect to HIV and AIDS, vaccine safety, and others likewise engage in consensus devaluation (Kirkby, 2008; Bauer, 2013). The motivation for those who find themselves amongst or sympathetic to a fringe community well outside of the scientific mainstream to attempt to discredit consensus is of course self-evident. It is also frequently accompanied by conspiratorial claims of intimidation and persecution at the hands of those within the consensus, as Dr. Tol recently demonstrated in front of the U.S. Congress (Crowther, 2005; Bauer, 2013; Tol, 2014b).

Error
4

Cites paper that misrepresents C13

“ T14 Intro: “Legates et al. tried and failed to replicate part of Cook's abstract ratings, showing that their definitions were inconsistently applied.” ”

L13 apply the technique of “impossible expectation”, one of the five characteristics of science denialism

Legates et al. 2013 (L13) inconsistently applies the definitions provided in C13. In addition, L13 misrepresents C13 by fabricating a category definition (catastrophist definition) that was not adopted in C13. L13 applies the technique of “impossible expectation”, one of the five characteristics of science denialism (Diethelm and McKee, 2009), to derive their argument that only 0.3% of the papers analysed in C13 endorsed the consensus. To arrive at this value, L13 raises the standard of endorsement of consensus to explicitly quantifying the human contribution to more than half of global warming, ruling out thousands of abstracts that explicitly or implicitly endorse AGW. In short, L13 derives its result by inappropriately ignoring the 3,833 abstracts explicitly or implicitly endorsing AGW.

L13 also fails to evaluate the consensus percentage for their revised abstract ratings. Instead, it simply evaluates the percentage of explicit AGW endorsements with quantification compared to the entire literature sample, including ‘no position’ abstracts. Thus it includes abstracts irrelevant to the question of the cause of global warming, and therefore do not accurately test for consensus. Consequently, L13 does not demonstrate that C13's definitions were inconsistently applied. Rather, it misrepresents and distorts the results presented in C13.

**Error
5**

Misrepresents stolen, private correspondence about training period

“ T14 Sec. 2: “Raters were not independent.” ”

T14 uses as a basis for this argument an excerpt from stolen private forum discussions (Lacatena, 2014) which is quoted out of context. Discussion of the methodology of categorising abstract text formed part of the training period in the initial stages of the rating period. When presented to raters, abstracts were selected at random from a sample size of 12,464. Hence for all practical purposes, each rating session was independent from other rating sessions. While a few example abstracts were discussed for the purposes of rater training and clarification of category parameters, the ratings and raters were otherwise independent. This was discussed in C13;

“While criteria for determining ratings were defined prior to the rating period, some clarifications and amendments were required as specific situations presented themselves.”

Independence of the raters was important to identify uncertainties based on interpretation of the rating criteria, but had little bearing on the final conclusion. Indeed, the conclusion is strengthened by the fact that the vast majority of rater disagreements were between no position and endorsement categories; very few affected the rejection bin.

T14 uses as a basis for this argument an excerpt from stolen private forum discussions which is quoted out of context

**Error
6**

Conflates literature analysis with survey of human subjects

“ T14 Sec. 2: “Unusually, Cook and his co-authors all rated abstracts”. ”

This procedure is not unusual – Oreskes and her colleagues rated all abstracts in Oreskes (2004). T14 confuses a survey of human subjects, in which it is unusual for the authors to participate in the survey, with an analysis of literature. In this situation, the subjects (abstracts) cannot be influenced by those conducting the survey.

T14 confuses a survey of human subjects, in which it is unusual for the authors to participate in the survey, with an analysis of literature. In this situation, the subjects (abstracts) cannot be influenced by those conducting the survey.

“ T14 Sec. 2: “In deviation from best practice (Mohler et al. 2008), no survey protocol was published; it is therefore not known whether the 4th rating was an ad hoc addition, which would invalidate the result.” ”

Mohler et al. (2008) discuss methodologies for surveying people as subjects. In that situation, survey protocols are designed to prevent contamination of subjects (survey participants). In the methodology of C13, raters play the role of interviewers while the abstracts act as the “subjects”. The purpose of survey protocol is to prevent the contamination of survey participants. The content of abstracts does not change no matter how often they are consulted. T14’s application of Mohler et al. (2008) to the methodology of C13 is inappropriate and irrelevant.



Misrepresents data required to replicate C13

“ T14 Sec. 2: “More information – specifically, rater ID, time of rating, survey protocol, and lab notes – was requested in vain, in contrast to best practice³ and journal policy⁴. John Cook refused to run diagnostic tests on the withheld data.”

T14 Sec. 4: “The full data-set would shed further light on possible causes of these problems but is unavailable. Cook has refused to release such diagnostic tests as the ratings profiles of individual raters, and the histogram of times between ratings.” and

T14 Sec. 5: “Instead, they gave further cause to those who believe that climate researchers are secretive (as data were held back)” ”

The release of privacy-protected identifying data discussed in T14 is unnecessary to replicate the C13 survey, and was withheld to protect the privacy of raters who were guaranteed anonymity.

The release of privacy-protected identifying data discussed in T14 is unnecessary to replicate the C13 survey, and the data was withheld to protect the privacy of raters who were guaranteed anonymity.

Timestamps for the ratings were not collected, and the information would be irrelevant. Two timestamps would be needed for each rating: rating-started and rating-ended. Moreover, the time to complete an abstract rating is dependent upon several factors such as the length of the abstract, technical level of the abstract language, and interruptions occurring during the rating. Hence T14 is incorrect to state that this information (which does not exist) would shed further light on C13.

All data relating to C13 of any scientific value was published at <http://sks.to/data> in 2013. Furthermore, the public were actively encouraged to replicate C13's research, with the launch of interactive webpage enabling people to rate climate papers and compare their ratings to C13's results (Cook, 2013).

The only data withheld was information that might be used to identify the individual research participants. This protocol was in accordance with University ethical approval specifying that the identity of participants should remain confidential and was approved by the publisher, Environmental Research Letters. This legal position has been maintained by the University of Queensland given its obligations under its research ethics policy (University of Queensland, 2014).



Conflates 'climate change' & 'global climate change' research

“ T14 Sec. 3.1: “In current usage, “climate change” means “global climate change”, unless otherwise specified.”

T14 Sec. 4.0: “The sampled papers are not representative of larger samples of papers, and probably not representative of the population either.” ”

The assumption by T14 that “climate change means global climate change” is at variance with many papers using the term “climate change” to refer to regional, not global, climate change (e.g., White et al., 1997; Gong and Ho, 2002; Meredith, 2005; Seneviratne et al., 2006; Maurer, 2007). This is self-evident from the admission that climate change is sometimes otherwise specified not to refer to global climate change. Consequently, the conclusion of T14 Section 3.1 that “global climate change” papers are not representative of “climate change” papers is irrelevant.

However, it is an interesting idea to compare the disciplinary distributions of Thomson Reuters Web of Science (WoS) search topics “global climate change” and “climate change”. T14 apparently also included the term “global warming” when doing this comparison (finding 53,359 papers), which complicates the analysis of T14. In our study “global warming” accounted for about 75% of our total records, compared to only about 25% for “global climate change” (some records are found with both search terms).

Unfortunately, because each paper can have multiple category assignments, little meaningful information can be derived solely by comparing the distribution of categories. Repeating these two searches in WoS we find that 3663 records (103% of the total retrieved) in searching “global climate change” are classified into one or more of ten top research areas (see table); 72,071 (109%) of the “climate change” records fit those research areas. Numbers over 100% indicate that the vast majority of records are categorised under at least one of these ten research areas, and some must fit two or more. The only research areas that are retrieved in WoS by the “climate change” search but not by “global climate change” are nutrition dietetics, philosophy, religion, virology, nursing, telecommunications, literature, and mining mineral processing.

The assumption by T14 that “climate change means global climate change” is at variance with many papers using the term “climate change” to refer to regional, not global, climate change

WoS Research Areas	GCC records	% of 3566 GCC	CC records	% of 47728 CC
ENVIRONMENTAL SCIENCES	1412	39.596	17904	37.513
ECOLOGY				
METEOROLOGY ATMOSPHERIC SCIENCES	417	11.694	17904	15.863
GEOLOGY	413	11.582	17904	17.522
ENGINEERING	234	6.562	2755	5.772
AGRICULTURE	213	5.973	2528	5.297
BIODIVERSITY CONSERVATION	205	5.749	2334	4.890
MARINE FRESHWATER BIOLOGY	200	5.609	2820	5.908
SCIENCE TECHNOLOGY OTHER TOPICS	199	5.580	2167	4.540
PLANT SCIENCES	187	5.244	1855	3.887
PHYSICAL GEOGRAPHY	183	5.132	3900	8.171

Table 1: Data records from WoS categorised by “research topic” from topic searches “global climate change” (3566 records) and “climate change” (47728 records). Columns show number and percent of total in each category. Searched May 24, 2014.



Fails to substantiate claims of bias

“ T14 Sec. 3.1: “The narrower query undersamples papers in meteorology (by 0.7%), geosciences (2.9%), physical geography (1.9%) and oceanography (0.4%), disciplines that are particularly relevant to the causes of climate change. This likely introduces a bias against endorsement.” ”

Simplistic assumptions about consensus likelihoods amongst certain disciplines, and a lack of consideration for the rectifying effect of removal of papers that are not climate-related, the proposed bias against endorsement is not supported or substantiated.

T14 proposes bias against endorsement based on the assumption that the papers published within a research area outside of climate-related expertise are less likely to endorse the consensus on anthropogenic global warming than are papers from within the field. This assumption is not substantiated. An alternative narrative may be that experts outside of the climate science field will defer to the expert consensus in their writing.

The effect that oversampling or undersampling has on the final consensus figure depends on the rate at which papers from each discipline were climate-related. Given the small over/undersampling biases reported by T14 (0.4% to 2.9%), simplistic assumptions about consensus likelihoods amongst certain disciplines, and a lack of consideration for the rectifying effect of removal of papers that are not climate-related, the proposed bias against endorsement is not supported or substantiated. Moreover, T14 does not offer any valid justification for why its preferred sample parameters are more representative of the climate literature than those in C13. T14 has simply shown that one possible set of sample parameters returns a slightly different proportion of papers in various climate-related fields than a second possible set of parameters.

**Error
10**

Fails to substantiate assumption regarding Scopus

“ T14 Sec. 3.1, on Scopus: “Earth and planetary sciences, the most relevant papers, are oversampled. This introduces a bias towards endorsement.” ”

Comparing disciplinary categories of Web of Science and Scopus is problematic. The categories in each system are defined very differently, with Scopus having particularly odd classifications (Jacsó, 2011). T14 offers no description as to how the broader Web of Science categories were aggregated to match those of Scopus. Furthermore, in both systems a single paper may have several category designations, rendering simple statistical comparison tests moot. T14 suggests that the Scopus category “Earth and planetary sciences” is oversampled by WoS compared to a larger set of records in that category from Scopus. However, that category doesn’t exist in WoS; thus the conclusion that it’s oversampled is based on the author’s undocumented assignment of WoS records to Scopus categories.

The conclusion that Web of Science (WoS) is biased towards endorsement of consensus is based on four assumptions:

1. Scopus includes younger journals than WoS
2. Scopus includes more obscure journals than WoS
3. Younger and more obscure journals are more likely to contain contrarian papers
4. Scopus contains more specialised papers than WoS

No evidence is supplied for any of these four assumptions and the conclusion of bias is therefore not demonstrated. T14 seems to suggest that although C13 was the most comprehensive analysis of scientific literature on global climate change of its kind, it was still not large enough. The larger sample size reduces the risk for biases in counting rare events. For example, Oreskes (2004) surveyed 928 abstracts and found no papers rejecting anthropogenic global warming, whereas the larger C13 set found several unequivocal examples.

T14 seems to suggest that although C13 was the most comprehensive analysis of scientific literature on global climate change of its kind, it was still not large enough

**Error
11**

Fails to substantiate claims of unrepresentativeness

“ T14 Sec 3.1: “Overall, though, Cook et al. both undersample and oversample papers that are likely to endorse anthropogenic climate change. Their sample is unrepresentative, but the direction of the bias is unknown.” ”

Evidence that any alternative search would change our conclusions is unconvincing, untested and unsubstantiated in T14.

T14 suggests that changing the search terms or database would give a more representative sample of the literature. We agree that no perfect database of the literature exists; both Web of Science and Scopus continuously add journals, change search options and refine their proprietary algorithms. Evidence that any alternative search would change our conclusions is unconvincing, untested and unsubstantiated in T14.

**Error
12**

False claims about Web of Science meta-data

“ T14 Sec 3.1: “WoS only considers the title, abstract and keywords, whereas the former [Scopus] uses meta-data too.” ”

Both the Web of Science and Scopus use meta-data generated by proprietary algorithms.

Both the Web of Science and Scopus use meta-data generated by proprietary algorithms for ‘topic’ (WoS) or ‘title-abstract-keyword’ (Scopus) searches (Testa, 2011, Jascó, 2011). Scopus lacks cited reference data for many records before 1996, and therefore earlier searches based on meta-data may be deficient (Jascó, 2011).

**Error
13**

Misrepresents private correspondence about “deja-vu”

“ T14 Sec. 3.2: “Fatigue may have been a problem,⁸ with low data quality as a result... Indeed, one of the raters, Andy S, worries about the “side-effect of reading hundreds of abstracts” on the quality of his ratings.” ”

T14 references private correspondence stolen during the hacking of a private discussion forum and posted on a blog. The correspondence was quoted out of context.

To justify this claim, T14 references private correspondence stolen during the hacking of a private discussion forum and posted on a blog. The correspondence was quoted out of context. The original comment refers to an impression that some abstracts were repeated in the database and that, after reading hundreds of abstracts, it was not clear whether that impression was real or not. This did not refer at all to fatigue. Raters were free to categorise abstracts at their leisure and were not subject to deadlines. In fact, rater experience and expertise grew along with the number of abstract ratings completed (See Error #15).

Error 14

Neglects to check consensus percentages

“ T14 Sec. 3.2: “Rolling averages are outside their 95% confidence interval far too often ... The results for skewness too indicate drift.”

T14 Sec. 4.0: “Furthermore, the rating data show inexplicable patterns” ”

The assertion of rater drift is based on analysis of the average endorsement level, using ordinal labels from 1 to 7. However, average endorsement level is not an appropriate statistic for making inferences about consensus percentages. C14 replicates T14’s analysis using the more appropriate consensus percentage calculated for 50-, 100- and 500-abstract windows. We find no evidence of the claimed rater drift. Consensus among initial ratings in a window falls outside the 95% confidence interval 2.8%, 3.2% and 1.7% of the time for 50-, 100- and 500-abstract windows respectively.

Average endorsement level is not an appropriate statistic for making inferences about consensus percentages

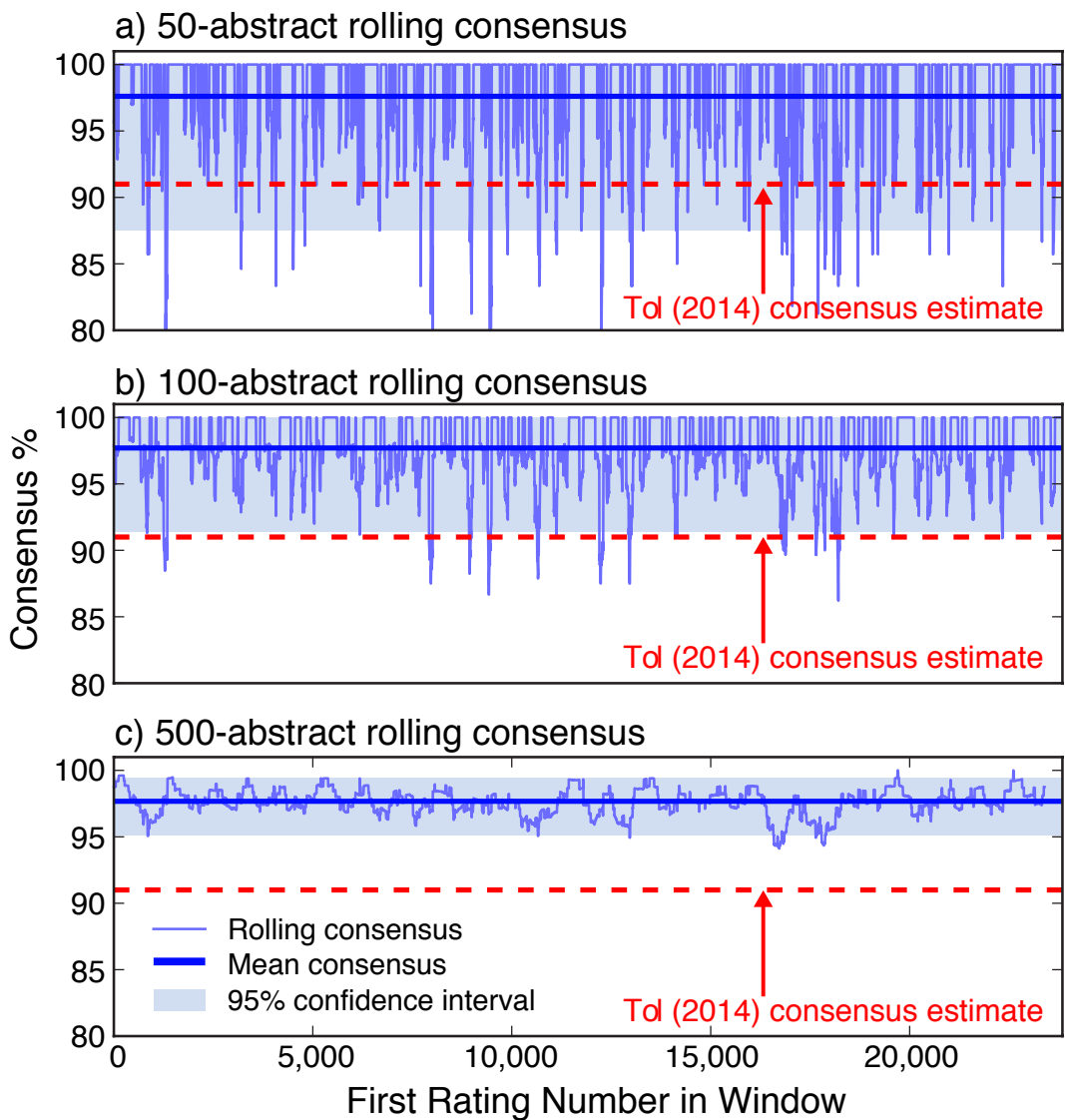


Figure 2. Calculated consensus (number of endorsements divided by total number that take a position) in rolling windows of 50- (a), 100- (b) and 500-abstract rolling windows (c) for the first 2 ratings of each paper prior to reconciliation. Shaded area indicates 95% confidence interval. Blue thick line indicates mean consensus. Red dashed line indicates T14’s recalculated 91% consensus.

The underlying problem is that the numerical labels attached to the categories are categorical and arbitrary. As noted by Healey (2011), categorical variables “are really just labels, not numbers at all. Statistics which require numerical variables are not appropriate, and often completely meaningless when used with non numerical variables” (similar comments may be found in most introductory statistics texts). While in some cases calculation of the mean of a categorical variable may yield informative results, this cannot be assumed (Knapp, 1990).

To illustrate this, consider two papers. Each paper is placed in one of 7 categories. The first paper is labeled an explicit endorsement with quantification (category 1), and the second is labelled as an implicit rejection (category 5). The mean of the category values is 3, equal to two papers which are both implicit endorsements (category 3). However the first pair of papers produce a consensus score of 50%, while the second pair produce a consensus score of 100%. The mean of the ordinal labels is not informative with respect to endorsement or rejection of the consensus.

The mean of the ordinal labels is not informative with respect to endorsement or rejection of the consensus.

**Error
15**

Contradicts cited experts on surveys

“ T14 Sec 3.2: “Fatigue may have been a problem,⁸ with low data quality as a result (Lyberg and Biemer 2008).”

”

According to T14's cited source, the effect of rating large numbers of abstracts would have the opposite effect to that stated in T14.

The fatigue discussed by Lyberg and Biemer (2008) addresses fatigue in survey subjects, describing how subjects taking repeated surveys can affect data quality. The abstracts, which are the subjects of the rating process in C13, cannot demonstrate fatigue. The raters performed the function of a survey interviewer in the process of rating abstracts. When contacted, Dr. Biemer confirmed that interviewers exhibit increased proficiency over time. According to T14's cited source, the effect of rating large numbers of abstracts would have the opposite effect to that stated in T14.

**Error
16**

Cites irrelevant, eliminated abstracts

“ T14 Sec. 3.3.1: “According to Cook, every abstract was rated twice; in fact, 33 abstracts were seen by only one rater.”

”

This statement is incorrect – 7 (not 33) abstracts were seen by only a single rater. These abstracts were observed upon first inspection to be non-peer-reviewed papers and immediately removed from the analysis after only one rating.

Error
17

Cites cherry-picking blog post

“ T14 Section 3.3.1: “The authors of the sampled papers also rated their work. Seven authors (including me) have disagreed with their papers’ rating.” ”

T14 once again references a blog post, discussing several scientific authors' self-ratings of their full papers. Author ratings are incorrectly compared to the phase of the research in which papers were categorised based on only the language in their abstracts.

Paper authors were invited to participate in a second phase of the research, in which they were asked to categorise their full papers based on statements contained therein regarding the cause of global warming. These ratings based on full papers are not directly comparable to those based only on abstracts. However, in both samples, 97% of abstracts or papers taking a position on the cause of global warming endorsed AGW. In the case of self-ratings, 1200 scientists rated their own papers. Among papers self-rated as stating a position on AGW, 97.2% endorsed the consensus. These results confirm the 97% expert consensus on AGW through two independent methods. T14 cites a blog that cherry picks 7 scientists.

1200 scientists rated their own papers. Among papers self-rated as stating a position on AGW, 97.2% endorsed the consensus.

Error
18

Cites report that falsely characterises C13 definitions

“ T14 Intro: “(Montford 2013) notes that Cook’s consensus is rather shallow – that carbon dioxide is a greenhouse gas, and that humans have played some role in observed climate change”. ”

Montford (2013) is a non-peer-reviewed report that falsely characterised the definitions used in C13, which precisely defined the levels of consensus for the study. Three definitions were used, with varying levels of stringency. These were 'humans are causing global warming' with papers minimising the human influence excluded, 'humans are causing global warming' (with the degree of contribution unquantified) and 'humans are the primary cause of global warming since 1951'. The consensus percentage among author self-rated papers for each definition was in the 96–98% range.

Montford (2013) is a non-peer-reviewed report that falsely characterised the definitions used in C13, which precisely defined the levels of consensus for the study.

**Error
19**

Incorrectly conflates abstract ratings with self-ratings

Author self-ratings of full papers are not a sub-sample of abstract ratings. They're an independent method to measure the consensus, measuring the level of endorsement stated in the full paper.

“ T14 Sec. 3.3.2: “Cook emphasizes that the consensus rates in the paper ratings and the abstract ratings are similar. A similar result in an unrepresentative subsample invalidates the finding.”

T14 Sec. 4.0: “Cook et al. failed to report that their data fail their own validation test.” ”

Author self-ratings of full papers are not a sub-sample of abstract ratings. They're an independent method to measure the consensus, measuring the level of endorsement stated in the full paper. This is a separate entity to the abstract rating, which measured the level of endorsement in the abstract text only. The difference between these two variables is discussed in C13.

**Error
20**

Fails to substantiate claim re categorising impact papers

Categorising each abstract based on its language is a far more precise and thorough procedure than making blanket assumptions and is the methodology adopted in C13. This suggestion from T14 would result in a less precise analysis.

“ T14 Sec. 3.4: “One could argue that impact papers should be rated as neutral or not at all ... 34.6% of papers that should have been rated as neutral were in fact rated as non-neutral .. Moving the papers on impacts and mitigation to 'neutral', 93% do not take a position. Correcting for misclassification, 95% of surveyed papers are silent on the hypothesis of anthropogenic climate change.” ”

Papers investigating the impacts or mitigation of global warming are doing so because they generally accept AGW implicitly (otherwise there is no reason to assume continued warming). Hence if any blanket assumption were to be made, climate impact papers should be categorised as endorsements. However, categorising each abstract based on its language is a far more precise and thorough procedure than making blanket assumptions and is the methodology adopted in C13. This suggestion from T14 would result in a less precise analysis.

**Error
21**

Fails to substantiate claim on implicit endorsements

“ T14 Sec. 3.4: “Implicit endorsements may be in the mind of the reader only.” ”

Implicit endorsements were based on the text published in the abstract, according to well-defined criteria in C13. T14's assertion also ignores the fact that paper authors also self-rated the endorsement of their own papers in the second phase of C13. Paper authors were instructed to rate their own papers based on the published text in their papers.

Implicit endorsements were based on the text published in the abstract, according to well-defined criteria in C13.

**Error
22**

False dichotomy between composition & endorsement

“ T14 Sec. 3.5: “The apparent trend in consensus is thus a trend in composition rather than in endorsement.” and T14 Sec. 4.0: “Cook et al. report a time trend towards greater endorsement. This, however, is due to an increase in the number of papers that are not on the causes of climate change.” ”

As consensus grows, research moves away from topics that provide additional evidence to support an accepted position and into new and challenging chapters of the story, such as impacts and mitigation (Shwed and Bearman, 2010). This evolution of research emphasis is discussed in C13. One might expect similar dynamics from an analysis of consensus, for example, on plate tectonics or evolution through natural selection. As the consensus continues to strengthen, the interest in directly addressing the underlying evidence supporting the consensus may diminish in favor of more novel questions, for example using the established evidence to inform policy-making and assess climate impacts.

Additionally, the compositional changes in the abstracts occur halfway through the survey period, but the consensus shift was observed in the first 25% of the period. The shift in composition and the shift in consensus do not coincide, thus negating T14's claim.

The shift in composition and the shift in consensus do not coincide, thus negating T14's claim.

**Error
23**

Irrelevant invocation of attribution research

“ T14 Sec. 4.0: “Theirs is not a consensus on the causes of climate change, but rather a vote of confidence by the broader climate research literature in the narrower literature on the attribution of climate change ... Researchers who think that climate change is real and anthropogenic are more likely to study climate impacts and climate policy than those who are unconvinced.”

This statement is factually incorrect. The consensus as defined in C13 is specifically regarding the causes of climate change. A separate survey to determine the consensus among climate attribution papers could be conducted. However, Tol has acknowledged,

“ “Published papers that seek to test what caused the climate change over the last century and half, almost unanimously find that humans played a dominant role.”

<http://andthentheresphysics.wordpress.com/2014/05/10/richard-tol-and-the-97-consensus-again/#comment-21162>

**Error
24**

False claim of polarisation

“ T14 Sec. 5.0: “Climate policy will not succeed unless it has broad societal support, at levels comparable to other public policies such as universal education or old-age support. Well-publicised but faulty analyses like the one by Cook et al. only help to further polarize the climate debate.”

This assertion runs counter to the conclusions of published scientific research. Public perception of scientific consensus on AGW is one of the strongest predictors of support for climate policy.

This assertion runs counter to the conclusions of published scientific research. Public perception of scientific consensus on AGW is one of the strongest predictors of support for climate policy (Ding et al., 2011; McCright et al., 2013). Rather than polarising, randomised experiments have found that presenting consensus information has a neutralising effect on both climate beliefs (Lewandowsky et al., 2013) and perception of scientific consensus (Kotcher et al., 2014). Corner et al. (2012) found little evidence for attitude polarisation in respect to climate change information; only that participants assimilated evidence in a biased manner.

References

- Abt, H. A. (2000). Do Important Papers Produce High Citation Counts? *Scientometrics*, 48(1), 65–70.
- Aklin, M., & Urpelainen, J. (2014). Perceptions of scientific dissent undermine public support for environmental policy. *Environmental Science & Policy*, 38, 173–177. doi:10.1016/j.envsci.2013.10.006.
- Anderegg, W. R. L., Prall, J. W., Harold, J., & Schneider, S. H. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 12107–12109.
- Bauer, H. (2013). "Denialism" – Who are the "denialists"? Skepticism about science and medicine. Retrieved from <http://scimedskptic.wordpress.com/2013/02/16/denialism-who-are-the-denialists/>
- Behe, M. (2009). Probability and Controversy: Response to Carl Zimmer and Joseph Thornton. *Evolution News and Views*. Retrieved from http://www.evolutionnews.org/2009/10/probability_and_controversy_re027491.html
- Boesch, D. F. (1999). The role of science in ocean governance. *Ecological Economics*, 31(2), 189–198.
- Bray, D (2010). The scientific consensus of climate change revisited. *Environmental Science & Policy*, 13, 340–350. doi: 10.1016/j.envsci.2010.04.001.
- Buela-Casal, G., & Zych, I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals. *Psicothema*, 22(2), 270–276.
- C4ID. (2009). How the Scientific Consensus can hinder Science. Centre for Intelligent Design. Retrieved from https://www.c4id.org.uk/index.php?option=com_content&view=article&id=209%3Ahow-the-scientific-consensus-can-hinder-science
- Cook, J. (2013). Measure the climate consensus yourself with our Interactive Rating System. *Skeptical Science*. <http://www.skepticalscience.com/Measure-climate-consensus-yourself-with-Interactive-Rating-System.html> (accessed 26 May 2014).
- Cook, J., Nuccitelli, D., Green, S.A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., & Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2), 024024+.
- Cook, J., Nuccitelli, D., Skuce, A., Way, R., Jacobs, P., Painting, R., Honeycutt, R., Green, S.A. (in press). Reply to Comment on 'Quantifying the consensus on anthropogenic global warming in the scientific literature: a Reanalysis'. *Energy Policy*.
- Corner, A., Whitmarsh, L., & Xenias, D. (2012). Uncertainty, scepticism and attitudes towards climate change: biased assimilation and attitude polarisation. *Climatic change*, 114(3–4), 463–478.
- Crowther, R. (2005). Academic Persecution of Scientists and Scholars Researching Intelligent Design is a Dangerous and Growing Trend. *Evolution News and Views*. Retrieved from http://www.evolutionnews.org/2005/12/academic_persecution_of_scient_1001701.html
- Diethelm, P., & McKee, M. (2009). Denialism: what is it and how should scientists respond?. *The European Journal of Public Health*, 19(1), 2–4.
- Ding D, Maibach EW, Zhao X, Roser-Renouf C, Leiserowitz A (2011) Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nat Clim Chang* 1:462–466.
- Doran, P., & Zimmerman, M. (2009). Examining the scientific consensus on climate change. *Eos, Transactions American Geophysical Union*, 90, 22.
- Gong, D. Y., & Ho, C. H. (2002). The Siberian High and climate change over middle to high latitude Asia. *Theoretical and applied climatology*, 72(1–2), 1–9.
- Healey, J. F. (2011). *Statistics: A Tool for Social Research*. Cengage Learning.
- Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., ... & Suddleson, M. (2008). Eutrophication and harmful algal blooms: a scientific consensus. *Harmful algae*, 8(1), 3–13.
- Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) (1996). *Climate change 1995: the science of climate change – contribution of working group I to the second assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, UK.
- IPCC, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [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, United Kingdom and New York, NY, USA.
- IPCC, 2014: Summary for Policymakers. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, CB et al (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jacsó, P. (2011). The h-index, h-core citation rate and the bibliometric profile of the Web of Science database in three configurations. *Online Information Review*, 35(5), 821–833.
- Kim, K. M. (1994). *Explaining scientific consensus: the case of Mendelian genetics*. Guilford Press.
- Kirkby, D. (2008). Former UK Science Chief – Vaccines Cause Autism: "What More Evidence is Needed?" *Huffington Post*. Retrieved

- from http://www.huffingtonpost.com/david-kirby/former-uk-science-chief_b_146717.html
- Knapp, T. R. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research*, 39(2), 121–123.
- Knorr, K. D. (1975). The nature of scientific consensus and the case of the social sciences (pp. 227–256). Springer Netherlands.
- Kotcher, J., Meyers, T., Maibach, E., Leiserowitz, A. (2014). Correcting misperceptions about the scientific consensus on climate change: Exploring the role of providing an explanation for the erroneous belief. Accepted for presentation at the 2014 annual conference of the International Communication Association.
- Lacatena, B. (2014). A Hack By Any Other Name – Part 1. *Skeptical Science*. Available at <http://www.skepticalscience.com/hack-2012-1.html> (accessed 26 May 2014).
- Legates, D. R., Soon, W., & Briggs, W. M. (2013). Climate Consensus and 'Misinformation': A Rejoinder to Agnotology, Scientific Consensus, and the Teaching and Learning of Climate Change. *Science & Education*, 1–20.
- Lehrer, K., & Wagner, C. (1981). *Rational Consensus in Science and Society: A Philosophical and Mathematical Study* (1981 edition.). Dordrecht, Holland ; Boston : Hingham, Mass: Springer.
- Lewandowsky, S., Gilles, G. & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, 3, 399–404. 10.1038/10.1038/NCLIMATE1720.
- Maurer, E. P. (2007). Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios. *Climatic Change*, 82(3–4), 309–325.
- McCright, A. M., Dunlap, R. E., & Xiao, C. (2013). Perceived scientific agreement and support for government action on climate change in the USA. *Climatic Change*, 1–8.
- Meredith, M. P., & King, J. C. (2005). Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophysical Research Letters*, 32(19), L19604.
- Meyer, S. (2009). Pro-Darwin consensus doesn't rule out intelligent design. CNN. Retrieved from <http://www.cnn.com/2009/OPINION/11/23/meyer.intelligent.design/>
- Miller, B. (2013). When is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement. *Synthese*, 190(7): 1293–1316.
- Mohler, P. P., Pennell, B. E., & Hubbard, F. (2008). Survey documentation: Toward professional knowledge management in sample surveys. *International handbook of survey methodology*, 403–420 (Psychology Press, Abingdon).
- Mulkay, M. (1978). Consensus in science. *Social Science Information*, 17(1), 107–122.
- NCEAS (2001). Scientific consensus statement on marine reserves and marine protected areas. *Journal of International Wildlife Law & Policy*, 4(1), 87–90. doi:10.1080/13880290109353975.
- Oreskes N. (2004) Beyond the ivory tower. The scientific consensus on climate change. *Science*, 306:1686.
- Oreskes, N. (2007) *The Scientific Consensus on Climate Change: How Do We Know We're Not Wrong? Climate Change: What It Means for Us, Our Children, and Our Grandchildren*, MIT Press.
- Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land–atmosphere coupling and climate change in Europe. *Nature*, 443(7108), 205–209.
- Shwed, U., & Bearman, P. S. (2010). The temporal structure of scientific consensus formation. *American Sociological Review*, 75(6), 817–840.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. (ed). (2007). *Climate change 2007: The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. 4). Cambridge University Press. http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg1_report_the_physical_science_basis.htm
- Testa, J., "Metadata & full text library holdings," http://wokinfo.com/media/pdf/backfiles-metadata_essay.pdf (accessed 17 May 2014).
- Tol, R. (2014). Quantifying the consensus on anthropogenic global warming in the literature: a re-analysis. *Energy Policy*, in press.
- Tol, R. (2014a). IPCC again. Richard Tol: Occasional thoughts on all sorts. Retrieved from <http://richardtol.blogspot.com/2014/04/ipcc-again.html>
- Tol, R. (2014b). Oral Testimony. Full Committee Hearing – Examining the UN Intergovernmental Panel on Climate Change Process, U.S. House of Representatives, Committee on Science, Space, and Technology. Archived at <http://science.edgeboss.net/wmedia/science/sst2014/FC052914.wvx>
- Tol, R. (n.d.). @RichardTol. [Twitter Profile page]. Retrieved from <https://twitter.com/RichardTol>
- University of Queensland (2014). UQ and climate change research. University of Queensland. <http://www.uq.edu.au/news/article/2014/05/uq-and-climate-change-research> (accessed 26 May 2014).
- White, J. M., Ager, T. A., Adam, D. P., Leopold, E. B., Liu, G., Jette, H., & Schweger, C. E. (1997). An 18 million year record of vegetation and climate change in northwestern Canada and Alaska: tectonic and global climatic correlates. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 130(1), 293–306.