

# Introducing the Extended Sniff Test (EST): A Method for Recreational Hostile Fact Checking

by a Literal Banana  
University of Twitter

---

## Abstract

Fact checking of scientific claims by lay volunteers, also known as recreational hostile fact checking or community-based science policing, is a growing hobby. A method for the evaluation of scientific claims by scientifically literate non-expert investigators is presented. The Extended Sniff Test, performed after an initial sniff test, uses the methods of Double Contextualization, Noun Abuse Assessment, and lay literature review, in addition to traditional literature review. As a case study, a suspicious paper is subjected to the Extended Sniff Test, and fails. A quick guide to the Extended Sniff Test is provided.

---

## Introduction

Volunteer fact checking of scientific claims is a hobby with limited appeal, but the advent of free access to an enormous number of academic papers has been a boon for enthusiasts. Writing informally about poorly-founded scientific claims can establish common knowledge among the writer's circle of readers, even if the criticisms do not become common knowledge among the relevant academics. The main hope of volunteer fact checking is not to change how science is practiced, but to change how scientific claims are evaluated by lay audiences (including scientists in fields other than the topic of the claims). In this paper, the term "investigator" is used to denote recreational hostile fact checkers, while other words are used for academic scientists.

In this paper, a community-based science policing method, the Extended Sniff Test (EST), is presented and applied to a sample suspicious study. In particular, the method focuses on a common but sadly underreported form of abuse, namely word abuse.

---

## Methods

A paper (Block and Block, 2006) was chosen for Extended Sniff Testing based on the subjective instincts of the author, who first encountered the paper while browsing the Wikipedia page of one of the paper's authors ("Jack Block", n.d.).

An initial sniff test was performed on the Wikipedia summary of the claim in the article ("Jack Block," n.d.), strongly suggesting the greater-than-trace presence of bullshit:

One of Block's studies drew particular notice in the news media. Published in *The Journal of Research in Personality* in 2006, it found that subjects who at 3 years old had seemed thin-skinned, rigid, inhibited and vulnerable tended at 23 to be political conservatives. On the other hand, 3-year-olds characterized as self-reliant, energetic, somewhat dominating and resilient were inclined to become liberals.

The article (Block and Block, 2006) was obtained and read.

## The Extended Sniff Test

The Extended Sniff Test (EST), introduced here, is a formalization of the practices the author uses when volunteer-fact-checking scientific claims. It is performed after an initial sniff is taken of the claim to be studied, if the sniff reveals the possible presence of bullshit. The EST contains three parts: Double Contextualization of the claim in question (DC), a Noun Abuse Assessment (NAA), and a lay literature review.

### Double Contextualization

General claims have often experienced traumatic amputation of their contexts. It is necessary to explore, with sensitivity, two types of context.

First, the allegedly generalizable claims of the study (its core claims) must be identified and mentally tested against the investigator's own available contexts. These contexts might include other scientific results, personal experience, and information from other times and places.

Second, the study's generalizable claim must be put back into the context of the study itself: the context in which the data were gathered must be reconstructed to the extent possible. The data gathering phase of the experiment is just as likely to be part of a science crime scene, spatiotemporally speaking, as the data analysis phase.

In terms of data analysis, numbers can be victims of abuse as much as words. The investigator should consider what the author chooses to report and not to report, particularly information about the *distribution* of the data. When an author gives limited or conflicting information about a distribution, the possibility must be entertained that important facts about the distribution are being obscured which might, if known, ruin the magic trick. For instance, an author may obscure the fact that a small set of extreme outliers is driving the result (see, e.g., Banana 2020, regarding Lewis Okun). A sense of the underlying distribution is a part of the context that an investigator might hope to recover.

## The Noun Abuse Assessment

Ironically, in order to perform a Noun Abuse Assessment (NAA), it is necessary to *not* believe the nouns. Words often deny abuse even when they are suffering greatly at the hands of an abuser. It is the responsibility of the investigator not to assume that a word carries the meaning that the investigator perceives it to contain in its context.

A word is being abused if ordinary intelligent readers would take it to have a particular meaning, but upon investigation, the meaning is found to be missing or changed.

Every word important to the conclusions of the study must be assessed. Abstract nouns are the most frequent victims of abuse, hence the name of the assessment, but other parts of speech also experience abuse. This is particularly true for verbs, whose hidden ambiguity is often exploited, and adjectives, particularly the adjectives describing human traits. The investigator should assess every abstract noun and human trait adjective in the title and abstract of the study for signs of abuse, and assess other words according to discretion and intuition.

The most common form of word abuse is when the authors use a survey instrument to measure a word, and then employ the word as a term for the instrument's output. The instrument often has little to do with the victimized word's rich and nuanced ordinary meaning (for example, the five factors in the Five Factor Model of personality). In the worst cases, these instruments are magic tricks, used to produce desired but unlikely results because, and not in spite of, their flaws.

But words may also be abused in the instruments themselves. Words must be assessed for *hidden ambiguity*: could they mean something different to a survey-taker/subject than what they are purported by the author to mean? If any of the important words in the study were measured using one or more survey instruments (e.g. scales, inventories, question sets), all instruments must be examined, and their words assessed for abuse. For example, words used on human abuse assessments, such as "kick" and "bite," do not at first glance appear to be victims of word abuse. But when human survey respondents are asked

about the words, they often reveal that the seemingly-unambiguous words were used to represent not violent kicks or bites, but playful acts performed and experienced in contexts other than abuse (Lehrner and Allen, 2014; see Banana 2020 for a full discussion). These human respondents may not have been abused, but the words were.

Words are usually abused so that the author may generalize inappropriately. Survey instruments are the most common method, because they are cheap and easy, and contain an often-hidden extra layer of words which may be further abused (the contents of the instrument itself). But survey instruments are not the only method for the violent amputation and grotesque surgical replacement of a word's meaning. The results of any research protocol, from games played in a laboratory to heart rate measurements, can be generalized by renaming them with a common word rich with everyday meaning, usually an abstract noun.

### Lay Literature Review

Publishing in academic journals is costly, in time and effort, and often actual money. Within the institutions of academia, criticizing another researcher's work can be costly in terms of future enmity. Criticizing a politically appealing result may also be costly in terms of reputation and apparent political loyalty.

Bloggers, however, are under fewer constraints. They often write anonymously, but even when they write under their real names, it is less costly in effort and reputation to criticize scientific claims outside the literature itself. A traditional literature review is not enough for EST; a lay literature review is essential in any case. In the lay literature review, the investigator searches the paper's title, author's names, and core claims (ideally in various paraphrases) on the internet, looking for blog posts and other media critical of the study's claims or methodology. If they exist, they often offer valuable shortcuts to discovering the nature of the paper's science magic trick.

A traditional literature review is still important, particularly any studies citing the target study. Volunteer investigators should pay particular attention to studies that allegedly confirm the target paper's result, and to the degree to which

they *actually* replicate the original study's methods and results. Often, small samples studied under relaxed epistemic conditions yield exciting results, but larger studies reveal the illusion (see, for an example, Jung 2010).

---

## Results

When contextualized broadly, both the Wikipedia summary of the claim and the reported results of the study strongly indicated the presence of bullshit. While the Wikipedia summary is provocative, the actual claims of the study are astounding.

Block and Block (2006) performed a longitudinal study of three-year-old children, in preschools in Berkeley and Oakland, California, between 1969 and 1970. The children were assessed by their teachers for traits on a list of 100 traits, the California Child Q-Set (CCQ).

When the children grew up and were 23, around 1990, 95 of them were tracked down and again assessed, by trained evaluators using the California Adult Q-Set (CAQ). Separately, subjects took seven surveys of various aspects of their political beliefs, and a composite score of six of these survey instruments was used to determine if they were politically liberal or conservative. It was noted that the adjectives "liberal" and "conservative" would need to be evaluated for word abuse using NAA, reported below.

The study's conclusions were reported in the form of a list of phrases. Of the 100 possible traits on the CCQ and CAQ, any that correlated with the composite political construct LIB/CON, with a p-value of under 0.1 (zero point one), were included and reported. This methodology includes many practices that are frowned upon today, and may be classified as p-hacking, among other issues (Wicherts et al., 2016).

Amazingly, the lists of traits for the boys who would be labeled CON on the LIB/CON scale were *exclusively negative* at age three, and almost exclusively negative at age 23 (the trait "Favors conservative values" is the only potentially neutral trait attributed to them). On the other hand, the boys who would be labeled LIB are

assigned *exclusively* positive traits at age 3, and overwhelmingly positive traits at age 23. For the girls, who are reported separately, more traits are assigned overall, and the girls who would be labeled CON are assigned mostly negative traits, while the girls who would be labeled LIB are assigned mostly positive traits. I reproduce the boys' lists here, along with the correlation sizes that Block and Block (2006) report. Positive correlations indicate LIB; negative indicate CON.

### “Conservative” boys

	Correlation with LIB/CON
<b>As assessed at age 3:</b>	
Is visibly deviant from peers	-.37
Appears to feel unworthy	-.35
Has a readiness to feel guilty	-.34
Anxious in unpredictable environment	-.30
Suspicious, distrustful of others	-.30
Tends to brood and ruminate or worry	-.29
Immobilized when under stress	-.27
Is easily offended	-.25

#### As assessed at age 23:

Favors conservative values	-.51
Uncomfortable with uncertainty	-.45
Behaves in a sex-typed manner	-.38
Judges self, others in conventional terms	-.37
Tends to proffer advice	-.32
Makes moral judgments	-.31
Compares self to others	-.28
Is power oriented	-.28

### “Liberal” boys

#### As assessed at age 3:

Is resourceful in initiating activities	.33
Seeks to be independent and autonomous	.32
Proud of accomplishments	.31
Is self-reliant, confident	.27
Becomes involved in what s/he does	.26

#### As assessed at age 23:

Is introspective, concerned w/self	.52
Concerned with philosophical problems	.44
Enjoys aesthetic impressions	.40
Has high degree of intellectual capacity	.39
Has insight to own needs, motives	.37
Genuinely values intellectual matters	.32
Complicates simple situations	.30
Had a wide range of interests	.28
Is an interesting, colorful person	.28
Tends to be rebellious, non-conforming	.27

These are enormous correlations, correlations that, if real, would presumably be noticeable by

an interested observer. The first stage of Double Contextualization requires that the investigator to imagine what the world would look like if the claims were true. Then the investigator must check if the world does in fact look that way.

Problems became apparent as the claim of the study was contextualized. Personal experience context added included the author's experience growing up among children in a rural, conservative area; the author's experience babysitting in conservative and liberal churches and playgroups; and the author's classroom experience with Catholic children in Bangalore, India, and kindergarteners in Cambridge, Massachusetts.

Scientific context was also added: two later, much larger studies (Fraley et al., 2012; Wegemer & Vandell, 2020) portray themselves as confirming Block & Block (2006). These studies examined different constructs from the Block & Block study, and achieved much more modest effect sizes for the few traits with reported correlations. Each study merits EST for its own claims, but this is outside the scope of the present study. (An investigator interested in these papers might focus in particular on the constructs designated “fear.” Emotion-related abstract nouns like “fear” and “disgust” experience a high rate of abuse.)

Some aspects of the context of the Block & Block study were retrievable, such as the context of the location and historical time. The kind of person who “Favors conservative values” despite growing up in Berkeley or Oakland in the 70s and 80s, an energetically liberal place and time, might not be generally representative of those holding conservative ideologies worldwide. Block & Block note that “relatively few” of the subjects were identified as conservative by the six-survey battery; the rarity of conservatives, or at least “conservatives,” might suggest that the Blocks were detecting a group who did not follow the prevailing norms. The few “conservative” boys were disproportionately labeled “deviant” at the age of three.

Suspiciously, Block & Block do not report much information about the *distribution* of political scores, as outputted by their six-survey composite, other than to say that there were few “conservatives” and they were more self-similar than the liberals. The authors describe the

distribution with a sort of word picture, rather than providing an actual picture.

An intensive (though by no means complete) search of the over 400 articles citing Block & Block (2006) was conducted using Google Scholar. Almost all viewed articles cite the study uncritically, including Fraley et al. (2012) and Wegemer & Vandell (2020). The present author identified only one article critical of Block & Block (2006), and the entire discussion of the study therein takes place over two paragraphs (Alford & Hibbing, 2007). No methodological issues are mentioned, but the authors appear to have performed a sniff test of some kind and found the study wanting. They do *contextualize* the study by noting that well-known conservative celebrities do not seem to possess the traits attributed to conservatives in Block & Block.

The paucity of criticism illustrates the need for a lay literature search. An internet search immediately turned up bloggers questioning the study's methodology in detail. One of these bloggers, Shawn Smith (n.d.), even reports on an email exchange with Jack Block<sup>1</sup> about the study's methods.

Smith identified severe word abuse in the Block & Block study. He found it not in the trait phrases, but in the LIB/CON composite measure, where the meaningful common words "liberal" and "conservative" are replaced with the output of a mostly-inscrutable composite of six out of seven administered surveys. He details major problems with all six of the scales used, and particularly with their combination. For example, one test consisted of a list of questions about political issues, but the study did not reproduce the questions. When Smith emailed Jack Block to ask about them, knowing that the exact wording of questions is a prime site of hidden word abuse, Block reported that he did not have access to the questions, and when asked if there was someone else who did, Smith reports that Block told him that the data were in storage at Harvard for at least two years, and could not be reached.

Smith located and read the other instruments used as well, as has the author, to the extent possible. One instrument was a five-option scale, from very liberal to very conservative, with no option for "don't know/other" (an option that Wegemer & Vandell, 2020, choose to include). Block & Block tested liberal and conservative traits independently, using two separate measures, but then undermined any value of this by smearing both measures back into the composite LIB/CON (Smith, n.d.). One of the scales appears to measure whether the subject is an angry buffoon, rather than political beliefs.

The correlations reported in Block & Block (2006), relaxed as the standards were for their inclusion, depend upon the LIB/CON construct as a valid measurement of political orientation if their conclusions are to be meaningful (Smith, n.d.). It appears that word abuse in the LIB/CON measure is the most powerful explanation for the magic trick performed here, although other factors also contribute.

In this Extended Sniff Test, Smith (n.d.) was enormously helpful in identifying word abuse, in this case adjective abuse. Sadly, many of the other contemporary blog posts are lost to time (see McNamara 2006 for broken links).

In summary, the balance of the evidence indicates severe noun abuse took place in the target study, particularly in the use of the LIB/CON scale of political orientation. Inappropriate generalization was detected. Overall, the study cannot be relied upon, and contributes nothing to the knowledge of human personality, development, and political affiliation. It may even be worse than nothing, since it is still widely cited, and dubious studies published as late as 2020 are still claiming to confirm its results (Wegemer & Vandell, 2020).

---

## Discussion

Jack Block was himself a major critic of the Five-Factor Model of personality (see, e.g., Block 2010). The author admires his advocacy for

---

<sup>1</sup> Jack Block died in 2010; the other study's author, his wife Jeanne Block, died in 1981, and her authorship is partially posthumous. It is never explained by the authors why the article was published sixteen years after the data were collected.

abused words. It is surprising that a scientist with Block's sophisticated understanding of the problems with the Five-Factor Model could engage in word abuse. It is important to recognize that detecting word abuse is a skill, and even those who understand the practice in principle are still potential perpetrators of word abuse, even if it is unwitting.

Jonathan Haidt (2012) cited the Block & Block study in *The Righteous Mind*, uncritically. Defending the study in a footnote (Haidt 2012, Chapter 12, Note 22), he said:

This study is widely misdescribed as showing that future conservatives had much less attractive personalities as young children. This seems to be true for the boys, but the list of traits for future liberal girls is quite mixed.

The reader may wish to review the lists of traits for the girls in Block & Block (2006) and decide if this is an adequate defense. Other authors cite

the Block & Block study with even greater enthusiasm; Haidt is only mentioned here because he is a famous public figure, whose scholarship the reader may already have some impression of. The Block & Block study is also cited uncritically in a 2019 paper on which Haidt is a co-author (Waytz et al., 2019). By the present author's count using Google Scholar, seventeen academic papers have cited Block & Block (2006) just in the first six months of 2020.

"It is an easy vice to generalise," as Sir Francis Galton said in 1884. One of the most common forms of unwarranted generalization is the substitution of meaningful common words with survey instruments and laboratory protocols that purport to measure them.

---

## Conclusion

The Extended Sniff Test may be a useful tool for recreational hostile fact checking and community-based science policing, particularly in the social sciences. The EST was able to detect problems with a 2006 study, while the field of social psychology as a whole was not able to detect such problems. If the EST or its concepts were to become common knowledge, it is possible that much abuse of innocent words could be prevented.

---

## References

Alford, J. R., & Hibbing, J. R. (2007). Personal, interpersonal, and political temperaments. *The Annals of the American Academy of Political and Social Science*, 614(1), 196-212.

A literal banana (2020). Ignorance, a skilled practice. <https://carcinisation.com/2020/01/27/ignorance-a-skilled-practice/>

Block, J. (2010). The five-factor framing of personality and beyond: Some ruminations. *Psychological Inquiry*, 21(1), 2-25.

Block, J., & Block, J. H. (2006). Nursery school personality and political orientation two decades

### EXTENDED SNIFF TEST (EST) QUICK GUIDE

Identify the core claims of the study

Double Contextualization:

- Consider known contexts and whether the claims are true of them
- Investigate the contexts in which the claims were produced

Assess for word abuse:

- all abstract nouns in the title, abstract, and central claims of the study
- all human descriptor terms (even when used to describe mice)
- all nouns, adjectives, and verbs in the title and central claim
- all words in all survey instruments used in the study

Look for words being given a different meaning from their usual meaning, usually using survey instruments or laboratory protocols to stand in for the ordinary meaning.

Review the lay literature.

Don't believe the words.

- later. *Journal of Research in Personality*, 40(5), 734-749.
- Fraley, R. C., Griffin, B. N., Belsky, J., & Roisman, G. I. (2012). Developmental antecedents of political ideology: A longitudinal investigation from birth to age 18 years. *Psychological science*, 23(11), 1425-1431.
- Galton, F. (1884). Measurement of character. *Fortnightly*, 36(212), 179-185.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- “Jack Block.” (n.d.) Wikipedia. Retrieved June 22, 2020. [https://en.wikipedia.org/wiki/Jack\\_Block](https://en.wikipedia.org/wiki/Jack_Block)
- Jung, C. G. (2010). Synchronicity: An acausal connecting principle. (From Vol. 8. of the collected works of CG Jung) (New in Paper). Princeton University Press.
- Lehrner, A. and Allen, N. (2014) “Construct validity of the Conflict Tactics Scales: A mixed-method investigation of women’s intimate partner violence” (*Psychology of Violence*, 4(4), 477–490).
- McNamara, M. (2006). “‘Cry baby’ study has critics bawling.” Retrieved on June 23, 2020. <https://www.cbsnews.com/news/cry-baby-study-has-blogs-bawling/>
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, 10, 813.
- Smith, S. T. (n.d.) “How to spot a broken study: The baby conservative project.” Retrieved on June 23, 2020. <https://docsmith.co/2007/02/how-to-spot-a-broken-study-the-baby-conservative-project/>
- Waytz, A., Iyer, R., Young, L., Haidt, J., & Graham, J. (2019). Ideological differences in the expanse of the moral circle. *Nature communications*, 10(1), 1-12.
- Wegemer, C. M., & Vandell, D. L. (2020). Parenting, temperament, and attachment security as antecedents of political orientation: Longitudinal evidence from early childhood to age 26. *Developmental Psychology*.
- Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in psychology*, 7, 1832.