DARPA-SN-17-57 (SCORE) Proposal (Round 2)

Abstract: This paper proposes **Ockham.io**, an open-source, web-platform, to automate in whole or part the algorithmic verification of scientific theories, hypotheses, and/or studies within the social and behavioral sciences per **TA3** - **DARPA-SN-17-57**.

Briefly, the approach taken herein involves: (1) explicitly **formalizing** embedded algebraic structures for computer verification, (2) **soft verification** through *corroborating reputability* of researchers, institutions, journals, and citations; (3) **natural language processing** to identify key terms, experimental variables, and concepts under study; (4) verifying **sound experimental design** and checking for **logical consistency**, **numeric error**, **bias**, and mathematical rigor; and (5) identifying how well the results, hypotheses, data, or conclusions **cohere** or are compatible with other high-credence theories.

Adam InTae Gerard rev. 1.1.2 - 7.28.18

Table of Contents

[DARPA-SN-17-57](#) (SCORE) Proposal (Round 2)

Adam InTae Gerard[1]

1. *Problema*

Bona Fide scientific research involves adherence to the ***Scientific Method*** – a general guideline for conducting and verifying empirical (using the five senses) research. The scientific method involves at least:

[1] Proposing an empirically testable hypothesis.

[2] Constructing a valid test methodology according to sound experimental design practices.

[3] Expressing clearly and unambiguously that method and hypothesis using formal (mathematical) models.

[4] Testing that hypothesis.

[5] Validating ones finding through experimental variable analysis.

[6] Submitting [1]-[5] above for peer review and replication (by following the same steps, the results can be experimental repeated in legitimate settings by other reputable scientists).

According to Popper, any theory that fails to exhibit [1] and [4] (which together express the ***Falsifiability Criterion***) are pseudo-science. Failures of [3] have generally led philosophers of science to be untrusting of the soft sciences. Any result failing [2] or [5] exhibits poor experimental design or analysis.

Presently, more than half of all research produced in Psychology fails [6].[2] The inability of such a staggering amount of

---

[1] I would like to thank many people who have contributed to working this proposal out anonymously, privately, and/or publicly and who thereby assisted in improving or clarifying this project. Where credit is due, it will be acknowledged upon request if it has not been included here (my apologies). Given the tentative and exploratory status of this proposal, some acknowledgements have been intentionally with-held until a future time (also per request).

[2] link and link

published research has acquired the infamous moniker dubbed the "Replication Crises" (Reproducibility Crisis).[3]

At best, the "Replication Crises" specifies some methodological error common to most present Psychological research. Identifying that specific issue could vastly improve the quality of all subsequent Psychological research.

At worst, the "Replication Crises" represents a systemic failure to comply with the scientific method resulting in wasted tax-payer dollars, poor public policy insights, and illegitimate medical research.

To be clear, the "Replication Crises" is not tantamount to the claim that "all and every piece of psychology research is made-up" but that, just as stated above, "over half of psychology research cannot be replicated."[4]

2. Solvency

Ockham.io is a proposed open-source, web-platform, to automate in whole or part the verification of scientific theories, hypotheses, or studies within the social and behavioral sciences to reliably and algorithmically determine the validity, credibility, or legitimacy of the same (per **TA3**).

True to its name, **Ockham.io** will do so primarily by using parsimonious algebraic structure validation, natural language processing (translating written studies into formal syntax), and simple equation validation of experimental variables and results.

[3] link and link

[4] The core logical concept of 'Some' but not 'All' (as employed in the basic machinery of ***First Order Logic***) represents a fundamental cognitive ability of rational agents. Here, we are clear that much of psychology meets the criteria laid out in **section 1**. For example, the DSM VI is strongly substantiated by neurological, anatomical, chemical, psychiatric, and biological findings throughout the cognitive sciences and medicine.

In other words, **Ockham.io** will allow scientists, academics, non-experts, families, and institutions to quickly and reliably determine whether a Psychology or Sociology study, theory, or hypothesis is trustworthy.

But **Ockham.io** will also assist in identifying the causes of the "Replication Crises", become an eventual archive of scientific data, assist in auditing scientific data (for errors), and standardize the use of mathematical models.

2.1. Outline

Ockham.io is an open-source **Application Programming Interface** (API) and **Web Application Service** supporting automated credibility assessment of social sciences research using machine learning algorithms.

The five core dimensions of assessment comprise:

[a] Explicitly **formalizing** embedded algebraic structures for computer verification.

[b] S**oft verification** through *corroborating reputability* of researchers, institutions, journals, and citations.

[c] N**atural language processing** to identify key terms, experimental variables, and concepts under study.

[d] Verifying **sound experimental design** and checking for **logical consistency**, **numeric error**, **bias**, and mathematical rigor.

[e] Identifying how well the results, hypotheses, data, or conclusions **cohere** or are compatible with other high-credence theories.

3. Algebraic Structures and Formalizing Experimental Terms

This stage consists in explicitly **formalizing** embedded algebraic structures (suitable to model theoretic treatments of scientific theories and theory classification per Suppes[5]) for computer verification.

Briefly, mathematical theories involve the five following components:

[**LANG**]	A formal language *L* with operators, logical symbols, and well-formed formulae (wff) specified by the grammar of *L*.
[**THRY**]	A set of wff, *S*, of a formal language *L*.
[**AX**]	A set of tautologies from which true sentences can be proven (bounded by the ***Second Incompleteness Theorem***[6]).
[**PROOF**]	Rules of inference such as *Modus Ponens*.
[**SEM**]	At least one structure which makes all the sentences of *S* true. This is alternatively called the model class or semantics for *S*.

3.1. Examples

An example of this approach can be found in the heavily condensed set-theoretic structural representation of Newtonian Mechanics which can be succinctly expressed as "$\langle P, \vec{s}, m, \vec{f}, \vec{g} \rangle$, where P is the set of 'particles', $\vec{s}$ is the position function, m is the mass function, $\vec{f}$ stands for the internal forces and $\vec{g}$ represents the external force function - all of them obeying certain postulates."[7]

[5] See da Costa 2008 pp. 5 and Suppes 2002.

[6]

[7] See Krause and Bueno 2007 pp. 3.

The basic edifice of ***Probability Theory*** can be expressed succinctly by way of a *finite probability space* - a triple $\langle \Omega, f, p \rangle$ where:

[1] Ω is a non-empty set of outcomes.
[2] f is a set of events such that:
  [a] $\Omega \subset f$.
  [b] f is closed under union and complementation.
[3] p is a function satisfying the following constraints:
  [a] $p: f \rightarrow [0,1]$.
  [b] $p\ (\varnothing) = 0$.
  [c] $p\ (\Omega) = 1$.
  [d] $p\ (E_1 \vee E_2) = p\ (E_1) + p\ (E_2) - p\ (E_1 \wedge E_2)$

3.2. Working Details

This stage accomplishes three primary aims: (1) to ensure that the terms defined are not semantically vague or ambiguous[8], operationalized to the extent that they are practically useful, and genuine; (2) specification of the underlying formal mathematical models to ensure the correctness of results[9]; and (3) to assist in accomplishing the other stages to follow.

Consider this randomly selected finding from a randomly selected psychology article:

(S) "Performance was better for the elaborative-interrogation group than for the control group (76% versus 69%), even after controlling for prior knowledge and verbal ability."[10]

---

[8] Imprecision in theoretical terms means that in two different instances we might be talking about two very different things despite using the same word. In such cases, the utility of the terms of used is highly dubious as are any results that might be derived from them.
[9] See Suppes 2002 pp. 30-34 for a brief description of this method.
[10] Randomly from Dunlosky, John, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathen, and Daniel T. Willingham 2013. [5] See Halvorson 2016.

The sentence above is a declarative sentence – e.g. a grammatically correct, meaningful, assertion expressing a state of affair. Such a sentence is exactly the kind of sentence that predicate logics were originally invented to capture. Translation schemes of that sort are very familiar – Carnap[5], Prolog, and the wider view called the Syntactic View of scientific theories[11] come to mind. Indeed, in *model theory*, a theory (such as a scientific theory or hypothesis) is defined as a set of grammatically correct sentences in some logic L. After defining our predicates and constants in a formal logic:

(1) $E =_{df}$ Constant | elaborative-interrogation group | $E \in \Omega$

(2) $C =_{df}$ Constant | control group | $C \in \Omega$

(3) $P =_{df}$ Function Predicate | Performance | $\Omega \in [0,1]$

We can express the declarative sentence S above as:

(*) $P(C) < P(E)$ or (**) $P(C) = .69\ \&\ P(E) = .76$

While simple, it should suffice to demonstrate how such a translation scheme is usually employed. Here, we go one step further and follow Suppes in terms of applying our First-Order expression above to the construction of a set-theoretic structure suitable as a model. This gives rise to a compact object we can use throughout the remainder of the stages.[12]

One potential boon, most social science research is conducted within a well-understand set of mathematical tools and structures.[13] That by itself significantly simplifies our project. Such structures canvass such topics as ***Number Theory***, ***Probability Theory***, ***Stochastics***, and ***Bayesian Logic***.

3.3. Natural Language Processing Specifics

[11] See Appendix and Halvorson 2016.
[12] See Suppes 2002 and "The Set-Theoretic Conception Of Science" 2007 for an example of such an approach.
[13] See Suppes 2002 pp. 129.

Ockham.io partly supports conversion (serialization) between fixed formats or file-types (such as but not limited to *PostScript*, *Markdown*, *HTML*, *LaTex*, *Adobe Portable Document Format*, etc.) with full support for all filetypes common to academic and scientific literature.

Given that scientific research is declarative (descriptive) in nature (and not imperative or optative) we can thankfully restrict the focus of **Ockham.io** to rich, declarative, **Natural Language Processing**.

Simple put, the basic elements that compose declarative sentences in English can be expressed succinctly:

[**DET**] Det → *the*, *a*, …

[**NOUN**] N → *man*, *ball*, …

[**AUX**] Aux → *will*, *can*, …

[**VERB**] V → *hit*, *see*, …

And, combine according to the following ***Generative Grammar***:

[1] S → NP + VP

[2] P → V + NP

[3] NP → Det + N

[4] Verb → Aux + V

These sentence formation rules are natively supported in all major commercial **Natural Language Processing** libraries such as *Carnegie Mellon's Sphinx* or *Stanford's NLP libraries*.

4. Soft Verification

Ockham.io will also perform s**oft verification** through *corroborating reputability* of researchers, institutions, journals, and citations providing a statistical correlation between various stake-holders (involved with funding and conducting scientific research) and credibility outcomes.

To be clear we are not committing the *fallacy* (an error of reasoning, logic, or cognition) of *Appealing to Authority* here.

Briefly reprised that fallacy formally involves the following inferential sequence:

(1) Person *P* is or is perceived to be an (epistemic) authority or expert about some topic *X*.
(2) *P* says that a hypothesis, study, or theory *T* within the topic *X* is true or correct.
(C) Therefore, *T* is true or correct.

It's important to note that the connection between (1), (2), and the inference to (C) above is a *necessary* connection. In other words, independent of evidence, the inference above asserts, just because someone is perceived to be an authority (which they might not be) and just because that person says something is true makes it true.

Instead, we will leverage external markers of credibility (a statistical concept) to assist in determining the *credibility* (or *probability* of truth) of a theory, hypothesis, or study.

4.1. Preceding Assessment

In general, reputable, or prestigious institutions create positive feedback loops to regulate and check the reputability and prestige of the studies commissioned by them. The reputation and prestige of that institution largely derives from the quality, novelty, and accuracy of the work that institution has produced.

4.2. Follow-Up Assessments

5.0 Natural Language Processing: Terms

Ockham.io supports natural language processing to identify key terms, experimental variables, and concepts under study.

This accomplishes or can help to accomplish several major or related objectives:

[1] To in part build a scientific lexicon or encyclopedia of terms (each of which in turn can be scored according to credibility)

[2] To help establish the credibility of the theory, hypothesis, or study Π by determining the usage of those terms within the written or spoken media items in which Π appears.

[3] To provide independent and useful metrics for other kinds of meta-scientific analysis.

On [2] above, we must be careful to separate mere reference or allusion to a non-credible term from the use of such a term as a *theoretical term*.

We therefore leverage natural language parsers to tease out specific markers that should or do modify our credence levels appropriately with respect to the works within which they are used (supported by the method outlined in **section 3.3**).

Consider the following examples that illustrate three different occasions of *use* and *manners of use*:

(a) "Aether was a silly concept…"
(b) "We propose Aether as an explanation…"
(c) Aether

(c) is the term, (a) is a mere allusion, (b) is an assertion of a theory, hypothesis, or study – i.e. a *theoretical term*.

We should be wary of proposals that rely on Aether as part of the proposed research. However, we should ignore sentences that merely refer to the concept of Aether.

Regarding [3] above, basic text-search pattern matching (using *Regex*, for example) is sufficient. Words that do not yet exist in stored vocabulary libraries will be flagged for inspection – these terms are most likely (indeed are almost exclusively likely) to give rise to formalizable *Predicates*, *Operators*, and *Constants* within a Formal Logic and simultaneously support the goals outlines in **section 3.3.**

6.0. Logic and Consistency Checking

Making explicit the underlying logic to determine logical consistency.

A minimum requirement should be that whatever logic is embedded into whatever mathematical theory that is determined per **2.1.** is logically consistent with respect to its own logical axioms, inference rules, and semantics.

Quantum Logic undergirds ***Quantum Mechanics***, ***Bayesian Logic*** is the foundation for much of ***Probability Theory***, etc.

Contradictions or inconsistencies discovered in this process will dramatically decrease the credibility of the theory, hypothesis, and study. Examples of the kinds of items that will be flagged as contradictions:

1. Any numerical error (inconsistency with the ***Theory of Numbers***).

2. Outright contradictions (e.g. P(x) = .4 & P(x) = -.4).
3. Inconsistencies in data or method (e.g. P(x) = .4 & P(x) = .8).

Given the use of appropriate equations (and numeric intervals for correlation strength) when performing statistical analysis, most if not all statistical validation issues would appear as a logical consistency error.

6.1. Statistical Validation

An additional constraint to improve the accuracy of the proposed system involves validating the key **experimental design features**[14]:

1. A sufficiently varied control group.
2. Sufficiently large sample size.
3. The existence of a control group.
4. Identification of experimental design flaws including biases in experimental design, high variability, sampling bias, selection bias, and so on.

Failure to satisfactorily obey those four basic constraints would result in a significantly reduced credibility score.

Standard ***experimental variable analysis*** (extreme deviation from the mean, statistically anomalous occurrence, null hypothesis, correlative strength, etc.) can be performed more simply since such analysis involves "internal" calculations between other experimental variables using sum of squares and covariance analysis.

6.2. Data-Dependent Validation

[14] Special thanks to Pawel Ngei of X-Team for drawing special attention to this point.

Identification of ***experimental design flaws*** sometimes relies on background data to adjudicate and assess the validity of the experiments or surveys conducted. To determine whether a study effectively mirrors the population at large (say to determine the presence of sampling bias), we must have access to accurate demographic information about the population.

Terms like `population` are context-sensitive (meaning that the intended reference is dependent on conversational context – in some occasions of use it can refer to a city's population, in others to a nation's, etc.) but are usually clearly disambiguated by the presence of adjacent adjectives or modifying clauses (e.g. – U.S. population, the population of Iowa, etc.).

Given a sufficiently formalized mathematical edifice and an effective formal translation scheme, each of these terms can be checked against an external data-set. Identified experimental design flaws can be used to assess the overall credibility of the study in question.

7.0. Coherence With Other (High-Credence) Theories

Ockhamio.io identifies *how well the theory, study, or hypothesis coheres with other high-credence theories, studies, or hypotheses*.

Given the construction of compact set-theoretic objects outlined in **section 3.**, it is possible to see how well the assertions of one theory, hypothesis, or study coheres with the assertions of other theories, hypotheses, or studies.

While coherence as a *sole foundation for justification* is dubious due to so-called impossibility results[15], coherence is is widely used as an additional assessment filter to check or

[15] See Meijs and Douven 2007 for an introduction to the debate and replies.

verify scientific discoveries[16]. The experimental compatibility of results, suppositions made, and predictions supported by two experiments or between two theories in part determines the credibility of research.

A finding that purports to overturn the (established) Theory of Relativity is exactly the kind of finding that warrants additional scrutiny, validation, and licenses increased concerns about credibility. Such a result may be proven true after all but only through such enhanced inspections.

In this way, tensions between theories, hypotheses, or studies can be identified and credence levels can be modified in accordance with the level of compatibility a theory, hypothesis, or study has with the (up-to) the rest and more interestingly with other high-credence theories, hypotheses, or studies.

8.0. Dataset Insights

A dataset combining these different stages can then be constructed with the intent to (in full or in part) automate credence level settings for theories, hypotheses, and studies.

Such a data set can play numerous roles – it can serve as a baseline to test algorithms against and it can be used to improve each iteration of the testing.

There are at least two ways to approach comparing the accuracy of an algorithm against such a data-set:

[16] See Thagard 2007 pp 28. While the exact terms 'coheres' or 'coherence' are perhaps not explicitly said by many scientists, the concept that a theory, study, or hypothesis should be logically consistent with our other best current scientific theories is often mentioned though not in those exact words. Theories, hypotheses, or studies that are not logically consistent warrant additional scrutiny and reservation. Consider the recent claim that the speed light had been broken – see Condliffe 2012.

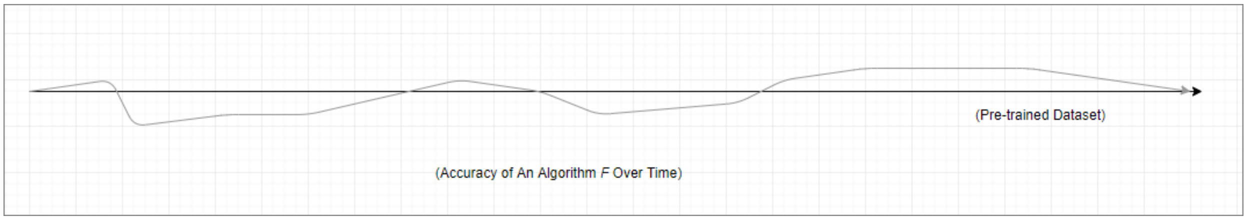

Fig. 1 – Algorithm Accuracy Approach Number One

Above, the data-set is generated before-hand. A disadvantage of that approach is that the replication credibility of the initial data set could change during development of the algorithm. An advantage is that this approach is likely easiest to implement.

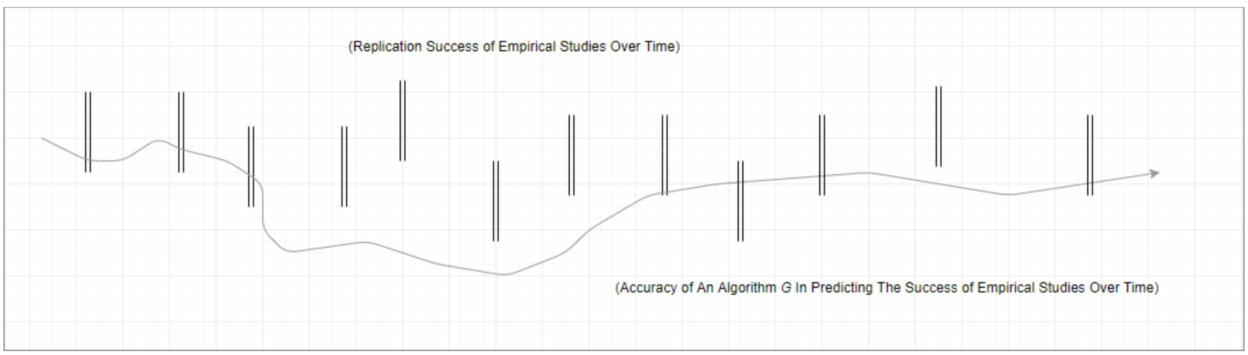

Fig. 2 – Algorithm Accuracy Approach Number Two

The second approach, above, demonstrates verifying an algorithm against future data. The algorithm is updated pending comparison against dynamically generated data.[17]

A mix of both approaches is encouraged here. Ultimately, if one of the two-approaches attains a high-degree of predictive power, it will solve for both kinds of scenarios.

9.0. Data Modeling

[17] This method is performed routinely in machine learning for price or stock prediction.

An early, simple, relational database implementation with data structures, models, and relations:

```
STRUCT_ID | DOMAIN | RELTIONS | LOGIC | FK_LIST_T_ID
1 | AE | F={1,3},G=DxD | CLASSICAL | 1,3,4

S_ID | TYPE | STRUCT_ID | CREDENCE | RESOURCE | NAME | FK_R_ID | FK_I_ID
1 | STUDY | 1 | .8 | https://... | "Studies of Normal People…" | 3 | 5

T_ID | TERM | CREDENCE | PREDICATE_CONSTANT
1 | AETHER | 0 | AE

R_ID | NAME | FK_ID_ID | CREDENCE
1 | Joe Everyman | 1 | .5

I_ID | INSTITUTION | CREDENCE
1 | Harvard | .9
```

Above we have five simple data models representing *mathematical structures*, *academic studies* or *resources*, *terms*, *authors*, and *institutions* (including schools, hospitals, and journals) along with several simple relationships defined between them.

Visually, those same data models are presented below:

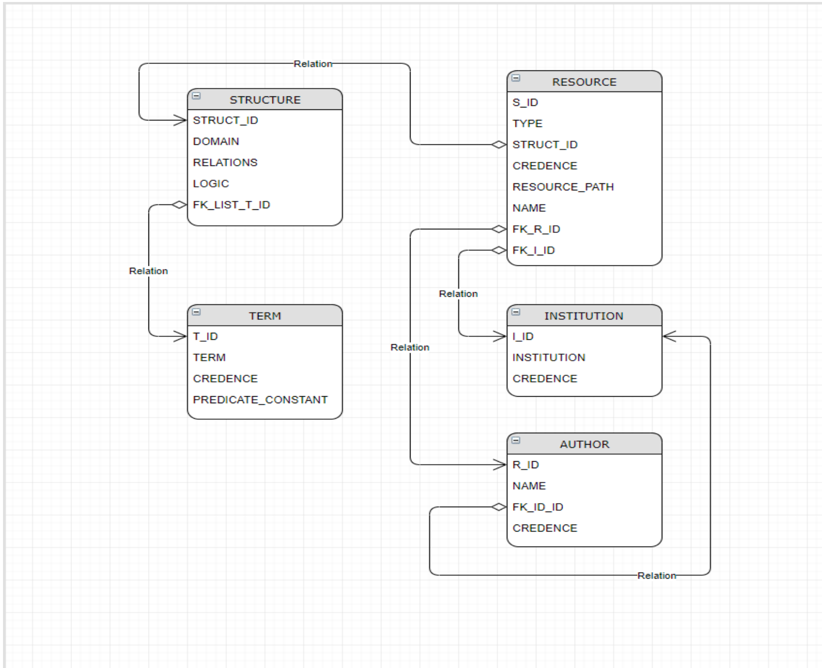

Fig. 3 – An early sketch of the domain architecture using an Entity Relation chart.

Fine-tuning (correcting, modifying) the data will be done manually for each *resulting stage*. Each of those fine-tuning operations will be audited for record-keeping and to assist in the automation of future fine-tuning operations.

Human intervention into correcting credence is encouraged until any errors in credence setting are ruled-out and enough accurate data would exist that the process could be largely automated with intervention being made on an as needed basis.

10.0. Overall Algorithm Specifics

The specific formula for calculating the credence of a specific theory, hypothesis, or study τ takes the form:

$$C(\tau) = \mathrm{COH}(\tau) \circ \mathrm{SV}(\tau) \circ \mathrm{T}(\tau) \circ \mathrm{LOGIC}(\tau)$$

Where each 'o' specifies an operation, 'C' denotes the overall credence function for τ, 'COH' denotes the to be specified coherence operation, 'SV' names the soft verification operation modifying the resultant credence by checking the credence levels of participant researchers and institutions, 'T' denotes any modification as a result of non-credible term use, and 'LOGIC' modifies the overall credence by checking logical consistency (presumably removing the credibility of the theory entirely if it isn't logically consistent).

Theories with little credibility could have a negative impact on the credibility of the individual researchers and the institutions that sponsored the work.

10.1. Credibility Markets and Schelling (Focal) Points

There are at least two broad classes of algorithms that can be introduced to assist in (semi-)automated verification of scientific theories. Below, I discuss two concepts that guide the overall algorithm design used in Ockham.io:

(1) ***Credibility Markets*** – so-called credibility markets regard the exchange of credibility-backed or social-standing-based goods (expertise, quality, credit-worthiness, asset rating, etc.).[18]

Similar work studying the impact of imposed policies and regulations on specific credibility markets has established a precedent that can be analogously implemented within the scope of verifying social science research. Here, specific credibility assignments and stage-based operations can be tested to determine their impact on the accuracy of the overall system.

[18] See Frenkel, Jacob and Guillermo A. Cavlo 1991.

In other words, we have a reusable template or approach to audit or make clear the impact of our credibility assignments at each stage (or substage) by mirroring the price and demand curves that result from analogous policies and regulations that impact credibility markets.

(2) ***Schelling (Focal) Points*** – a Schelling point refers to a coordination strategy in information-poor, partly secretive, or outcome-unknown circumstances (including jury voting, trust, street car traffic stops, credibility, etc.).[19]

Such focal points often involve consideration of the mirroring or coordination strategy itself used by agents within such games to overcome the lack of certainty.

Both systems can be implemented in several ways each of which can then solve to the correct algorithm (generating it dynamically) from some supplied data-set.

In the first case, each step of the multi-step approach laid out in the preceding sections can be tested against some control then compared against an independently scored data set to determine correctness of assigned credibility. This allows us to track the impact a single modification to a stage may have as discussed directly above.

In the second case, the coordination problem at hand involves correctly rewarding credible empirical research and punishing less credible research. *Replication* (or the ability for other scientists and researchers to recreate the results of an experiment) could be the *focal point itself* but we can also delimit this approach to a single sub-stage or stage and its ability to effectively mirror some subset of the overall data set.

A more concrete example of that would see several articles passed into a single stage of Ockham.io. Those that fail to

[19] See Janssen 2006.

reach a certain threshold are removed, each one that succeeds stays. Those that remain through all stages can be used to augment some of the missing algorithm specifics laid out in **Section 4.0**. We can study certain details at each stage that can help to filter or give a richer understanding about what kinds of articles, who wrote them, where they were published, what institutions published them, etc.

Another approach sees multiple algorithms being compared against each other for fitness with succeeding algorithms surviving and the less successful algorithms being weeded out over several epochs of testing. Here, again, the *Schelling Point* can play an essential role in helping to determine what the fitness cut-off point is.

One might be concerned about whether the legitimacy of the core algorithm must itself be credible. We think so, but its credibility is demonstrated by its correctly identifying the credibility of the others.

Both approaches can be undertaken *simultaneously* to increase the probability of success.

11.0. Open-Source, Blockchain, and Academia

It is the hope of this author that the dataset and overarching system above would be open-sourced and made publicly available for scrutiny, improvement, and to aid in public policy decision-making across many institutions.

Blockchain technologies represent a potential and exciting but non-essential way to implement **Ockham.io**'s core functionalities while making all data public through the blockchain ledger, distributing the data for redundancy, and leveraging distributed computing to perform the actual calculations (which are relatively minor).

Such a blockchain implementation might additional provide a self-sufficient source of funding, economic incentives for participation, etc.

12.0. Works Cited

Alston, William P. "Sellars and the 'Myth of the Given'." Philosophy and Phenomenological Research, 65 no. 1, July 2002. 〈https://www.jstor.org/stable/3071107〉.

da Costa, Newton C.A. "The mathematical role of time and space-time in classical physics." *Arxiv*, 7 February 2008. 〈https://arxiv.org/pdf/gr-qc/0102107.pdf〉.

Dunlosky, John, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathen, and Daniel T. Willingham. "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology." *Association for Psychological Science*, vol. 14, no. 1, 2013, pp. 4-58. 〈http://www.indiana.edu/~pcl/rgoldsto/courses/dunloskyimprovinglearning.pdf〉.

Condliffe, Jamie. "Did Scientists Really Just Break the Speed of Light?" Gizmodo. 7 May, 2012. 〈https://gizmodo.com/5908206/did-scientists-really-just-break-the-speed-of-light〉.

Frenkel, Jacob and Guillermo A. Cavlo. "Credit Markets, Credibility, and Economic Transformation." *Journal of Economic Perspectives* vol. 5, no. 4, 1991, pp. 139-144. 〈https://pubs.aeaweb.org/doi/pdf/10.1257/jep.5.4.139〉.

Halvorson, Hans. "Scientific Theories." *The Oxford Handbook of Philosophy of Science*. Paul Humphreys ed. Oxford: Oxford University Press, 2016.

⟨http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199368815.001.0001/oxfordhb-9780199368815-e-33⟩.

⟨http://philsci-archive.pitt.edu/11347/1/sci-theories.pdf⟩. (preprint)

Krause, Décio and Otávio Bueno. "Scientific Theories, Models and the Semantic Approach." *Principia* 11.2, 2007, pp. 187-201.

Janssen, Maarten C. "On The Strategic Use Of Focal Points In Bargaining Situations." Journal of Economic Psychology, vol. 27, no. 5, October 2006, pp. 622-634.

⟨https://www.sciencedirect.com/science/article/abs/pii/S0167487006000481⟩.

⟨https://papers.tinbergen.nl/06040.pdf⟩.

Meijs, Wouter and Igor Douven. "On the Alleged Impossibility of Coherence." *Synthese*, vol. 157, no. 3, 2007, pp. 347-360.

⟨https://ai2-s2pdfs.s3.amazonaws.com/8e51/dcf5747d7e1e20df03823e73303f9c2806b3.pdf⟩.

Suppes, Patrick. *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI, 2002.

⟨http://web.stanford.edu/group/cslipublications/cslipublications/pdf/1575863332.rissbook.pdf⟩.

Thagard, Paul. "Coherence, Truth, and Development of Scientific Knowledge." *Philosophy of Science*, 2007, vol. 74, pp. 28-47.

⟨http://cogsci.uwaterloo.ca/Articles/coherence.truth.pos.2007.pdf⟩.

"The Set-Theoretic Conception Of Science." *Scientific Progress. Synthese Library* vol. 153, 2007, pp. 108-122.

⟨https://link.springer.com/chapter/10.1007/978-1-4020-6354-1_11⟩.

13.0. Appendix

I. Key Terms

Epistemic Credence: the level of *trustworthiness* or probability that we think a proposition, hypothesis, declarative sentence, or theory ∏ is true. Tied to the level of justification we think that ∏ has.

Model Theory: the study of set-theoretic semantic structures that make all the sentences of formal language true.

Dataset: a set of mappings *D* between an input(s) and an output(s) such that given enough accurate information and time, reasonably accurate and consistent predications can be made by a neural network when trained on *D*.

Neural Network: an implementation of a biological neural network such that artificial neurons are connected into layers that are in turn connected to each other. Inputs are passed in to the first layer of neurons, functions applied to those inputs such that an output is produced by the last layer of the network.

Formalization: here, to *translate* or *transpile* a *fragment* of a *natural language* (up to and including the natural language itself) into a *formal language*.

Formal Language: A mathematically rigorous language used to, without ambiguity or vagueness, specify a domain (topic and scope of inquiry), artificial language, rules of inference, and truth-conditions. A *logic*, a *grammar*, and a *semantics*.

Nonmonotonic Inference: a type of non-classical inference whose consequent or conclusion is altered upon addition or deletion of information. *Abductive reasoning* a paragon example of this type of reasoning.

II. Sciences of the Psyche

Psychiatry: the medical practice of diagnosing, identifying, and treating mental illness by distinguishing underlying physiological, neurological, chemical, anatomical, or biological conditions, phenomena, or causes.

Psychology: the non-medical study of mental phenomena research and practice of which is usually limited to correlational studies and clinical counseling.

Neuroscience: the scientific study of the brain and its operation by recourse to biological, anatomical, computational, linguistic, and chemical explanations.

Cognitive Science: the application of computational, formal, mathematical, and linguistic methods to the study of mind, thought, and brain.

III. Key Relevant Philosophy Concepts

Semantic View of Scientific Theories: a scientific theory should be conceived as the *model class* of a set of grammatically valid sentences of some formal language.

Syntactic View of Scientific Theories: a scientific theory should be conceived as a set of grammatically valid sentences of some formal language.

Epistemic Scientific Structuralism: the position that our knowledge of the world is limited to structural knowledge of the world (while remaining silent about or denying the possibility of knowledge of *things-in-of-themselves*).

Connectionist Functionalism: *functionalism* is not to be confused with *symbolic computationalism*, *functionalism* is merely the thesis that (1) mental states are either identified by what they do rather than what they are made of (e.g. - some substance, *haeccity*, or primitive *thisness*) and/or (2) determined by the role they play or the system of which they are part (and hence, could be implemented in numerous substrates). The main version of this view is *connectionist* *functionalism* which identifies the functional systems giving to rise to mental phenomena with neural networks.

A priori: what can be known without experience – unknown theorems of mathematics are just such examples.

A posteriori: that which cannot be known *a priori* – e.g. Empirical science requiring observation of some physical phenomena – discovering a new species, a new exoplanet, or force of nature.

IV. Relevant Considerations

The utility of correlations: when is a correlation considered to demonstrate a robust relationship between phenomena? In other words, when does a real correlation obtain rather than a "fake", superfluous, or merely coincidental one?

Logical positivism and the problem of reducing scientific statements to sense data or observables:

There have been numerous issues pertaining to identifying or reducing high-level scientific terms to underlying primitive sense-datum or other observables. Some of these philosophical worries have bled into other fields under moniker "Symbol Grounding" (per the "Symbol Grounding" problem in Artificial Intelligence).

I do not have room here to discuss this too deeply but I will briefly say that there are at least three problematic assumptions interwoven into such concerns that give rise to the problem in the first place: (1) Noumena / Phenomena distinction and the picture of representation that arises from it[20], (2) reductivism grounded in object-based ontologies[21], and (3) the Cartesian dogma of inner 1st person mental experience and outer 3rd person reality.

This proposal is not reliant on dogmatic foundational terms and such considerations are at this point largely irrelevant.

Limitations of Artificial Intelligence and Machine Learning: it takes time to develop a robust dataset that can be trained on a neural network to reliably and accurately generate correct outputs.

V. Professional Qualifications

I have a humble background in philosophy and software.

For a full summary of my professional and academic history please look at my LinkedIn.

A resume has also been attached.

[20] Some of these worries fall away when metaphysics is recast along modern mathematical lines. Isomorphism (or even a weaker morphism), Univalent Foundations (which equates *identity* and *isomorphism* as a fundamental axiom), and non-objectual ontologies help to blur these two Kantian distinctions. There is not so much an inner and outer (mental and external world), per say, but rather two layers of roughly approximate structural data (forthcoming).

[21] Which necessitates that there are some "Given" empirical primitives (see Alston 2002) or atoms that all sensory experience is ultimately or rests upon foundationally.