# Response to Reviewer 2 Comments

| | | |
|---|---|---|
| **Open Review** | (x) | I would not like to sign my review report |
| | ( ) | I would like to sign my review report |
| English language and style | ( ) | Extensive editing of English language and style required |
| | ( ) | Moderate English changes required |
| | (x) | English language and style are fine/minor spell check required |
| | ( ) | I don't feel qualified to judge about the English language and style |

| | Yes | Can be improved | Must be improved | Not applicable |
|---|---|---|---|---|
| Does the introduction provide sufficient background and include all relevant references? | ( ) | ( ) | (x) | ( ) |
| Are all the cited references relevant to the research? | ( ) | ( ) | ( ) | ( ) |
| Is the research design appropriate? | ( ) | ( ) | ( ) | (x) |
| Are the methods adequately described? | (x) | ( ) | ( ) | ( ) |
| Are the results clearly presented? | (x) | ( ) | ( ) | ( ) |
| Are the conclusions supported by the results? | (x) | ( ) | ( ) | ( ) |

The authors initially present a way recursively to identify repetitive components in long strings, and show how to use this knowledge to compress the string, that is, to map the original string onto a shorter one, from which the contents of the initial string can be fully recovered. The authors argue that this can be used for structures more general than strings, but the way in which they perform such extensions is the straightforward mapping of these more general (say 2-dimensional) structures, on strings. The possible amount of compression can then be used to provide a measure of the information contained in the string. They then proceed to considerations of a more general nature concerning the origin of life and other matters of general interest.

**Response 1:** Thanks for the comments. The reviewer is absolutely right that in this manuscript, we basically just used strings as examples (also a set of 2D patterns, as the reviewer mentioned, to show a slightly more general case) to explain the "ladderpath" concept and related definitions. Yet, we also wrote a few paragraphs to discuss how the ladderpath approach can be applied to chemical molecules, by pre-specifying one examplified scheme of the possible generation-operations and the definition of the length unit of a ladderpath (line 717-748). Indeed, it's not a full-length discussion, but our motivation is to explain the ladderpath approach itself.

Thanks for the reviewer to point this out, but as it is a new approach and involves quite a few new concepts, we think sticking to the simplest cases (e.g., strings) and explaining it step by step would be a good choice, with a few slightly more general examples for extrapolation. In principle, the ladderpath approach can be applied to many types of entities.

The general considerations which end the paper appear to be of significant interest, but there is, in my view, an earlier issue which must be addressed: essentially the ``ladderpath'' algorithm described by the authors is nothing else than a variation on the LZ compression algorithm proposed in 1978 by A. Lempel and J. Ziv. The authors dismiss this with the remark: ``Zempel-Liv is related to the optimal rate of lossless compression and does not account for the hierarchical structure of a sequence''. This is questionable on two counts: first, the ladderpath algorithm developed by the authors also is of the lossless compression type, and second, it is simply wrong to state that the LZ algorithm does not take into account the hierarchical structure possibly present in the input. Indeed, the ``ladderpath'' algorithm is best viewed as some variant on LZ. There are technical differences, but these refer mainly to the fact that the LZ algorithm is fully specified, whereas the ``ladderpath'' algorithm is not.

**Response 2:** Thanks very much for the comments and for mentioning the "significant interest".

First of all, the sentence associated with "lossless compression" is not meant to criticize LZ algorithm, and indeed, the reviewer is right that the ladderpath approach can also be applied on strings as a lossless compression approach.

We admit that "LZ algorithm does not account for the hierarchical structure of a sequence" is an incorrect statement (and we thus deleted it). Our original intention is actually to say that revealing the hierarchical information is not the main motivation for LZ algorithm. Now we realized that, thanks for the reviewer pointing it out, it's exactly because the hierarchical relationships among substrings are naturally taken care of by the way the LZ algorithm handles the string (e.g. adding new items into the dictionary etc.), that the information carried by the original string can be efficiently compressed. We have revised the main texts and cited the references accordingly (some recommended by the reviewer) and also corrected some typos (line 74-78).

The motivation here is to show a proof-of-concept algorithm and code to calculate the shortest ladderpath for a small string (the idea is that if an algorithm for short strings can be specified, other algorithms for long strings, images, molecules, proteins, 3D objects etc could be developed and sophisticated in the future). As Lempel-Ziv is a highly efficient and widely used algorithm for string compression, our intention is not to develop a practical or even more efficient algorithm for string compression.

The reviewer is absolutely right that LZ is fully specified. But we hope that the algorithm to calculate the shortest ladderpath is also sufficiently explained: 1) Section 2.7 of the main text described the main logic of the algorithm; 2) Appendix C used an example to explain the algorithm in detail (we have added a paragraph in Appendix C to explain what do we mean by "pre-determined systematic slicing", to be clearer); 3) We provided the source code openly published on GitHub which is obviously fully specified, i.e., every step is clear. Since there could be many methods to calculate the shortest ladderpath (as mentioned in the main text), there are many code implementations. So in the main text, we first explained the general logic (with an example in Appendix C) and then showed a source code which is merely one implementation to calculate the shortest ladderpath, out of many.

Indeed, as Lempel-Ziv is also a lossless compression algorithm, it shares some similarities with this ladderpath algorithm in the aspects of general ideas of hierarchy and a few coding techniques. Nevertheless, besides their different motivations, the outputs of the two algorithms are also different. We have added two paragraphs in the main texts to explain this point and the point above (line 490-500).

To be clearer, here we compare the results given by the ladderpath algorithm (provided here and published on GitHub) and the Lempel-Ziv algorithm. As the former outputs a partially ordered multiset and the latter outputs a compressed string, they are hard to compare directly. Nevertheless, in both algorithms, the final results are actually determined by the calculated scheme of slicing the string, so we can compare their calculated slicing schemes. We use the example string "ABCDBCDBCDCDEFEF" (section 2.2 in the main text):

The slicing scheme of the shortest ladderpath (calculated from the algorithm provided) for this string is

"A | BCD | BCD | B | CD | C | D | EF | E | F"

corresponding to the ladderpath Eq (1) in section 2.3 in the main text;

The slicing scheme of the Lempel-Ziv algorithm for the string is

"A | B | C | D | BC | DB | CD | CDE | F | E | F"

We can see that they are different.

The authors now claim a relation between Shannon entropy for an isolated sequence and ladderpath compression. Such remarks are undoubtedly correct, but they do not appear to be new: the articles quoted at the end of this report all have something to say on this subject.

**Response 3:** Thanks very much for the comments.

In Discussion Subsection 4.4, we intended to discuss how ladderpath could help investigate the question "why is life ordered". It certianly relates to entropy, so we spent the first 1/3 of this subsection briefly discussing the connections between entropy and ladderpath. Indeed, in the aspect of string compression, the connections between Shannon entropy and compression algorithms such as Lempel-Ziv have been discussed (as in the references the reviewer suggested), e.g., entropy is always a lower bound for compression, and the Lempel-Ziv algorithm is asymptotically optimal, achieving this lower bound. We added a few sentences to refer to these points and cited the references accordingly.

Nonetheless, we focused more on the aspect of molecules and living systems, e.g. how to use ladderpath to describe 1 mol of butane at equilibrium, and what the connection is between the entropy and the ladderpath description. Surely it's not a full-length discussion, but our main point here is to raise the question "why life is ordered" and mention the potential connections with ladderpath, and that's what the subsequent 2/3 of this subsection focused on, with which we hope to convey the idea that ladderpath can provide a different angle when investigating this type of questions. Maybe the title of this subsection is a bit misleading, which made the reviewer wonder if we're talking about Shannon entropy at length. We thus changed the title, and also added a few sentences to make the intentions of this subsection clearer (line 751-759).

Finally, it is not clear to me to what extent the claim made in the Abstract is fulfilled in the paper: ``From the ladderpath two measures naturally emerge: the ladderpath-index and the order-index, which represent two axes of complexity." But it appears that these two measures simply add up to the total length of the string: surely, the 2 are thus not in any meaningful sense independent measures of complexity. The paper thus does not appear to yield anything else than the variation on the Shannon entropy for individual signals already discussed in the literature.

**Response 4:** Thanks very much for the comments.

As we always have Lambda(X) + Omega(X) = S(X) (by definition), indicating that the two axes of complexity are not independent for a particular target or target system X. It is true indeed, but in fact, any target can be placed in a particular position in these coordinates. This is because although the three indices are constrained by Lambda + Omega = S, two of them are free. Referring to Figure 2 in the main text, if we fix the size of a target, by rearranging the patterns of this target, it can move freely in the coordinates (e.g., imaging rearranging among pattern [i], [iii] and [vi]).

It is exactly because of the internal structure and information of this target, its coordinates are fixed. In fact, Lambda and S, or Omega and S could also be chosen as the axes, but our motivation here is to relate the intuition of complexity with the axes, and the intuition of complexity often comes from difficulty and order. So, we chose Lambda and Omega. We have added a paragraph in the main texts to make this point clearer (line 433-443).

As it stands, the paper does not adequately describe considerable previous work on this and related subjects. Further, it is not clear to this author that there is anything significantly new in the paper, which should therefore be rejected.

**Response 5:** Thanks for this comment. Besides having revised the parts associated with Lempel-Ziv, we also revised the introduction to introduce more related subjects and research (cited the references accordingly and some recommended by the reviewer). And we realized that maybe previously we have given the work the name "ladderpath theory" which is too ambitious, so we changed it to "ladderpath approach".

By this reply and the revision, we hope we could convince the reviewer that there is sufficient novelty in this manuscript, including the definition of ladderpath and how to calculate it specifically, the definitions of ladderpath-index and order-index, the ladderpath-system, the isolated system and non-isolated system which are linked to how to interpret unknown signals or messages, etc.

We've also corrected some typos and minor parts here and there.

Thank you very much for your time.


1. Jacob Ziv, IEEE Transactions on Information Theory, Vol.~IT-24, No.~4, JULY 1978, p.~405, Coding Theorems for Individual Sequences
2. Jacob Ziv and Abraham Lempel, IEEE Transactions on Information Theory, Vol.~IT-23, No.~3, MAY 1977, p.~337. A Universal Algorithm for Sequential Data Compression
3. Jacob Ziv, and Neri Merhav, IEEE Transactions on Information Theory, Vol.~39, No.~4, JULY 1993, p.~1270. A Measure of Relative Entropy Between Individual Sequences with Application to Universal Classification
4. Jacob Ziv and Abraham Lempel, IEEE Transactions on Information Theory, Vol.~IT-24, No.~5, September 1978, p.~530, Compression of lndividual Sequences via Variable-Rate Coding
5. Hansel G., Perrin D., Simon I. (1992) Compression and entropy. In: Finkel A., Jantzen M. (eds) STACS 92. STACS 1992. Lecture Notes in Computer Science, vol 577. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-55210-3_209