Three-way Decision, Three-World Conception, and Explainable AI

Yiyu Yao(✉)

Department of Computer Science, University of Regina, Regina, SK, Canada S4S 0A2
Yiyu.Yao@uregina.ca

Abstract. Three-way decision is about thinking, problem-solving, and computing in threes or through triads. By dividing a whole into three parts, by focusing on only three things, or by considering three basic ingredients, we may build a theory, a model, or a method that is simple-to-understand, easy-to-remember, and practical-to-use. This philosophy and practice of triadic thinking appears everywhere. In particular, there are a number of three-world or tri-world models in different fields and disciplines, where a complex system, a complicated issue, or an intricate concept is explained and understood in terms of three interrelated worlds, with each world enclosing a group of elements or representing a particular view. The main objective of this paper is to review and re-interpret various three-world conceptions through the lens of three-way decision. Three-world conceptions offer more insights into three-way decision with new viewpoints, methods, and modes. They can be used to construct easy-to-understand explanations in explainable artificial intelligence (XAI).

Keywords: Three-way decision · Three-world conception · Three-world model · Thinking in threes · Trilevel thinking · SMV space · Explainable AI

1 Introduction

With the ever-increasing power, functionality, and applications of intelligent machines and systems, the issue of the explainability takes center stage. The recent research trend in explainable artificial intelligence (XAI) suggests that a machine must effectively explain its internal processes and decisions, in order to gain human understanding, trust, and acceptance [2,12]. As a prerequisite for producing effective explanations, it is necessary to study human ways to

Y. Yao : I would like to express my thanks to Professor Duoqian Miao and Professor JingTao Yao for organizing the IRSS President's forum and for encouraging me to write this paper. I am grateful to the reviewers for their encouraging and constructive comments. This work was partially supported by a Discovery Grant from NSERC, Canada.

J. Yao et al. (Eds.): IJCRS 2022, LNAI 13633, pp. 39–53, 2022.
https://doi.org/10.1007/978-3-031-21244-4_4

perceive, think, and act. With an understanding of human ways to think, understand, and act, a machine may explain its processes and decisions by building a model aligned with human mental models. Driven by such motivation, this paper explores particular mental models, namely, three-way decision as thinking in threes, three-world conception as thinking through three worlds, and the relationships between the two, as well as their applications in XAI.

There are two related types of issues around the notion of an explanation.[1] One type concerns the meaning, functionality, and properties of the explanation, as well as various formal models of explanation. In the context of XAI, an intelligent machine explains its working processes and results for the purpose of facilitating human understanding and building human trust. In a wide context of scientific enquiry and discovery, one of the goals and tasks of science is to explain the world, i.e., to seek "mathematically formulated and experimentally validated impersonal principles that explain a wide variety of phenomena" [36]. The other type focuses on the communication of an explanation, involving the structures and the construction process of the explanation. To some degree, an appropriate structure plays a crucial role in constructing an easy-to-represent, easy-to-communicate, and easy-to-understand explanation. The focus of this paper is on the latter type of issues. By applying the principles of three-way decision, I discuss ways to construct and communicate explanations with triadic structures.

The rest of the paper is organized around three objectives. Section 2 provides an overview of a theory of three-way decision with the objective to establish a basis for this study. The objective of Sect. 3 is to introduce, in light of three-way decision, a framework for studying three-world conceptions, that is, thinking through three worlds. In particular, I examine three -world models. The objective of Sect. 4 is to outline a possible application of three-way decision and three-world conception in constructing human-friendly explanations in data science, human-machine co-intelligence, and explainable artificial intelligence (XAI).

2 An Overview of Three-Way Decision

In 2009, I introduced the concept of three-way decision (3WD) [39] to provide a semantically sound interpretation of the three types of decision rule (i.e., acceptance, rejection, and undecided) derived through Pawlak rough sets [18,19] and probabilistic rough sets [40]. Further studies have shown that three-way decision is a much richer concept, with wide-ranging applications. Since 2012, I have been refining a new theory of three-way decision, consisting of thinking, problem-solving, and computing in threes [41–43,45,46]. Three-way decision has fostered

[1] The two types presented here are related to the distinction, suggested by Achinstein [1], of an "explaining act" and an explanation as a "product" of an explaining act. Ruben [23] made a similar distinction through "process and product." The first type is more about an explanation itself. The second type relies on an understanding of an "explaining act" that includes both the formulation and the communication of an explanation.

new research areas, such as three-way classification, three-way clustering, three-way data analytics, three-way formal concept analysis, three-way approximations of fuzzy sets, three-way conflict analysis, three-way recommendation systems, three-way granular computing, and many others. The field has grown substantially since its inception, with researchers from around the world contributing to a significant number of papers, edited books, journal special issues, workshops, and special sessions on three-way decision. For the current state of research and development of the art, science, and practice of three-way decision, a reader may consult the reports by Yang and Li [37], Wei et al. [35], and Yao [38] based on networks analysis and bibliometrics analysis.

Thinking in threes (i.e., triads consisting of three things) or triadic thinking is perhaps one of the most common mental models, metaphors, and structures, such as a tripartite scheme, a three-part theory, a three-element structure, a three-pillar framework, a three-word slogan, a three-character story, a three-generation classification, a three-level architecture, a three-version design of a product, a third grey option in addition to commonly used dichotomies (e.g., Yes and No, black and white, good and bad, positive and negative), a third middle point through the balancing and synthesis of the two opposites, and many more [3,4,15,22,33,43,46]. We humans and particularly scientists have an intriguing preference for a ternary patterned theory, model, or explanation of reality [20]. As an illustration, we may give three examples of thinking in threes. The first example is building a model of explanation for explainable artificial intelligence (XAI) based on the What-Why-How triad[2]: What are the results? Why are the results meaningful? How are the results derived? The second example is the MIT Sloan Management Review's short podcast, Three Big Points[3], in which each episode presents a mold-breaking idea in ten minutes with three useful takeaways. The third example is the effective use of threes in writing a great paper[4]: the three C's of paper structure consisting of the Context for introduction, the Content for results, and the Conclusion for discussion; the ABC (Accurate, Brief, and Clear) of straightforward writing; the DEF (Declarative, Engaging, and Focused) for choosing a title. In particular, advice on straightforward writing is summarized in three sentences: "Never choose a long word when a short one will do. Use simple language to communicate your results. Always

[2] This example will be further examined in the later part of the paper. For an actual application, we may point at the earlier expert system MYCIN that uses the What-Why-How triad, in which an explanation subsystem focuses mainly on Why and How questions to justify the decision of the system or to educate the user [32]. The triad is equally useful for enhancing human intelligence and guiding human behavior [46]. For example, the Golden Circle leadership model, introduced by Sinek [29], is based on the Why-How-What triad, which advises that every organization and everyone of us should know the three most important things: why we do (i.e., purpose and goals), how we do, and what we do. The same Why-How-What triad was used by Clear [6] in his three-level model of behavior change, focusing on what we believes, what we do, and what we get.

[3] https://sloanreview.mit.edu/audio-series/three-big-points/, accessed May 20, 2022.

[4] https://www.nature.com/articles/d41586-019-01362-9, accessed May 20, 2022.

aim to distill your message down into the simplest sentence possible." We can find many examples that explore the power triads for crafting great, powerful, and memorable speeches [10].

These examples show that we do commonly build an argument, a model, or a theory by thinking in threes. To provide further supporting evidence, it may be more constructive by giving three good reasons why we humans think in threes. The first explanation is the cognitive basis. It has long been recognized that we humans can only hold up a few things in the short-term working memory [7,16]. While there does not exist a general agreement on the exact number, which may range from two to nine, three seems to be a pivoting one. Another related result is our subitizing ability to tell immediately, without counting, the number of items presented to us when the number of items is small, typically fewer than six [14]. This may explain why the very first three Roman numbers are written as one, two, and three vertical lines, respectively, the very first three Chinese numbers are written as one, two, and three horizontal lines, respectively, and the pattern breaks at and after the fourth number. The third result is our natural ability to form patterns in order to make sense of the reality and our experiences. Three seems to be the minimum number of things required to form a meaningful and useful pattern. Drawing from these results of human cognition, thinking in threes comes naturally and may be an innate capacity.

The second explanation is the evolutionary basis. From an evolutionary point of view, we are better at older skills than at newer skills. Counting a few things and thinking about a small number of things, as evidenced by the 'one, two, three, four, many' and 'one, two, many' types of numerical systems [8], may be older skills in the process of human evolution. We, in fact, learned counting and thinking in small numbers at a younger age. Thus, we excel at skills of thinking in small numbers. It may be argued that thinking in threes is one of the products of evolution or early childhood learning.

The third explanation is the cultural basis. The number three plays an essential role across many cultures [9,25]. The number three typically represents completeness, harmony, and perfection, as expressed by the following quotations [25]:

- All good things come in threes. (Folk saying)
- A threefold cord is not quickly broken. (Bible)
- All was divided into three. (Homer)
- A whole is that which has a beginning, middle and end. (Aristotle)
- The Triad is the form of the completion of all things. (Nichomachus of Gerasa)
- Three is the formula of all creation. (Honoré de Balzac)
- The One engenders the Two, the Two engenders the Three and the Three engenders all things. (Tao Te Ch′ing)

Using a triad of three things for perceiving, understanding, interpreting, and representing the reality seems to be a universal practice across different cultures. Triads are perhaps one of the most used structures when crafting a story, a speech, a theory, or a worldview. For example, Schneider [25] stated, "Whenever there are three, as the three knights, three musketeers, three wise men, or three wishes, there is *throughness*, rebirth, transformation, and success." To a large

extent, our cultural immersion experience further re-enforces an inclination and a preference towards thinking in threes.

Given the omnipresence of triadic thinking on the one hand and a lack of a formal theory on the other, a theory of three-way decision has been proposed and received much attention in recent years [41–43,46]. The theory is about a systematic study of thinking, problem-solving, and computing in threes. By attaching specific interpretations and meanings to various triads, we can obtain different models and modes of three-way decision. In the rest of this paper, I interpret a triad in terms of three worlds, which gives rise to thinking through three worlds.

3 Thinking Through Three Worlds

This section examines three triadic structures, namely, a Venn diagram of three sets, a triangle, and a concentric tricircle, for thinking through three worlds.

3.1 The Concept of Worlds

The concept of "the world" is perhaps one of the most commonly used notions or metaphors for us to describe, view, and understand the reality and our relationships to the reality. The word "world," particularly, 'the world,' is used in various contexts with multiple meanings [34]. According to Webel [34], "the world" is "a linguistic and historical construction" and "an abstraction, a concept, or idea." It is how the "meaning-creating organisms frame the boundaries of their being-in-this-world." The view of "world as idea" [26,34] provides a starting point for exploring how we use the concept of worlds to understand the reality and to guide our conducts, namely, how to observe the world, how to make sense of the world, and how to change the world.

We may categorize and characterize things into different worlds in many ways, for example, from a temporal, spatial, functional, positional, or contextual consideration. We typically divide various aspects of the reality, for example, a group of geographical regions, a timeline of developments, a discourse of discussion, a family of human activities, etc., into a number of different and interrelated worlds. By restricting to a particular world, we limit our investigation within that world in the context of other worlds. Conceptually, we can talk about the inside, the outside, and the boundary of a world, which offers three interpretations and understandings of the same world. By considering different worlds, we can make comparisons, study their interconnections and influences, and shift our attention by switching between different worlds. While a single world presents a local view, multiple worlds give rise to a global view.

Our extensive living experiences on the planet earth as "the world," our relentless search for a better world, and our constant cultivation of a superior inner world all suggest the value of "world as idea." Conceptualizing the reality in terms of different worlds leads to both intuitive and in-depth understandings. By combining the principles of three-way decision as thinking in threes and the

view of "world as idea," we immediately arrive at a paradigm of thinking through three worlds. There are abundant examples of three-world thinking. In the contexts of information processing, knowledge management, problem solving, and human experience, for example, we have:

- The three-world theory of the reality and knowledge by Popper [21], consisting of World 1 of physical objects, World 2 of mental activities, and World 3 of human-created things.
- The theory of three worlds of mathematics by Tall [31], consisting of conceptual embodiment, operational symbolism, and axiomatic formalism.
- The classification of three worlds of knowledge by Mouton [17], consisting of the worlds of everyday life (lay knowledge), science (scientific knowledge), and metascientific reflection (metascience).
- The theory of triadic game design by Harteveld [13] through balancing the three worlds of reality, meaning, and play.
- The theory of collective human experience by Shaw [28] in terms of the three worlds of commonsense, religion, and science.

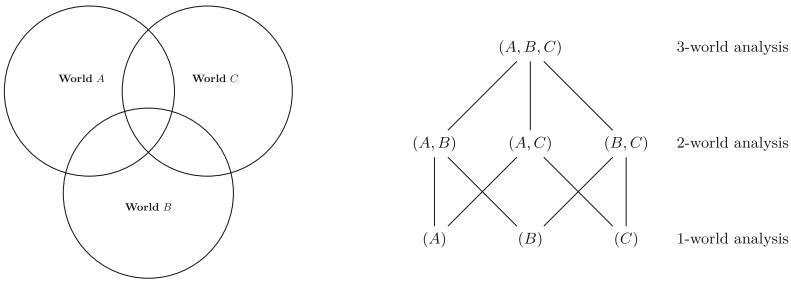
Other examples of three-world thinking in more general contexts include various triads, such as the material-intellectual-spiritual three worlds, the three worlds above-below-upon the earth (i.e., heaven, hell, and earth), the three worlds of yours-mine-theirs, etc.

It becomes evident that three-world thinking, with an understanding of "world as idea," offers a new direction for expanding the study of three-way decision as triadic thinking. In the rest of this section, I examine three particular models by organizing and arranging the three worlds in three different ways.

3.2 A Venn Diagram Model of Three Worlds

One methodology of the three-world view and analysis is to divide the discourse of discussion into three possibly overlapping and relatively independent worlds. There may exist multiple ways to construct three worlds. Any particular three-world configuration is only one of the many possible simplifications or representations of the reality. In general, the division between the three worlds is not a clear cut and some issues may appear in two or all three worlds. The Venn diagram in Fig. 1(a) depicts such a set-theoretic view of three-world thinking. Each world represents a particular view and focuses on some particular aspects. While a set covers issues in a world, the complement of the set covers issues not in the world. An intersection of two or three worlds represents their joint issues. With three worlds, the eight disjoint and possibly non-empty regions are, in terms of set intersection, $A \cap B \cap C$, $A \cap B \cap \bar{C}$, $A \cap \bar{B} \cap C$, $A \cap \bar{B} \cap \bar{C}$, $\bar{A} \cap B \cap C$, $\bar{A} \cap B \cap \bar{C}$, $\bar{A} \cap \bar{B} \cap C$, $\bar{A} \cap \bar{B} \cap \bar{C}$, where $\bar{A}$ denotes the set complement of A. In so doing, we can systematically investigate issues in the eight regions.

Alternatively, we may consider only regions constructed by using set intersection, representing issues in the overlapping regions of different worlds. In this



(a) Venn diagram of three worlds (b) Trilevel analysis

Fig. 1. Thinking through three worlds with a Venn diagram

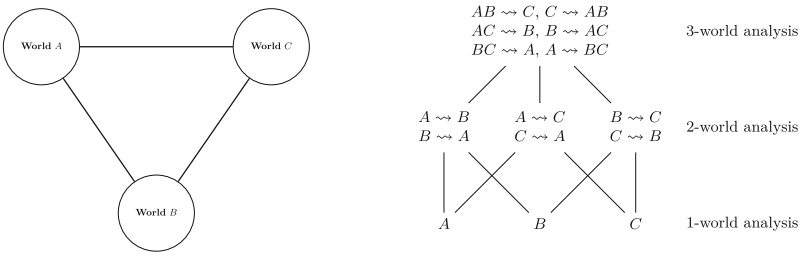
way, a three-world method offers a trilevel seven-element analysis in Fig. 1(b), where the comma corresponds to set intersection. The result is, in fact, a set-theoretic model of three-way decision [47]. The bottom level of 1-world analysis focuses on each world independently, the middle level of 2-world comparative analysis shifts attention to issues brought by interactions of two worlds, and the top level 3-world integrative analysis looks into more complicated interactions of three worlds. To have a holistic view, it is necessary to have investigations at the three levels, both individually and jointly.

Tall's [31] three-world model of mathematical thinking may be interpreted based on the Venn diagram of three worlds. While each individual world focuses on a particular type of mathematical methods and skills, a join of two worlds shifts the focus to the integration and combination of the respective methods and skills. Mouton's [17] classification of three worlds of knowledge and Shaw's [28] three-world theory of collective human experience may be similarly explained based on the Venn diagram of three worlds.

3.3 A Triangle Model of Three Worlds

For studying relationships, influences, and transformations of different worlds, a triangle of three worlds, given in Fig. 2(a), may be an appropriate configuration [46]. In the triangle, each world is linked with the other two worlds. Links between two worlds may have many different interpretations, for example, dependency, transformation, support, and others. In this way, a triangle may, in fact, offer various models. Figure 2(b) describes a model of trilevel analysis based on a triangle configuration of three worlds, where $\rightsquigarrow$ denotes support or transformation. We examine individual worlds at the bottom level, relationships between two different worlds at the middle level, and relationships among three worlds at the top level.

Popper's [21] three-world model of human knowing and knowledge is typically interpreted as a triangle. World 1 of physical objects exists first. Through World 2 of mental activities and processes, humans observe and make sense of World 1.



(a) Triangle of three worlds (b) Trilevel analysis

Fig. 2. Thinking through three worlds with a triangle

The results are human-created things that exist in World 3 as abstract ideas and/or in World 1 as physical objects. The things in World 3, created by World 2, may be used to change World 1. Humans are constantly searching for a better world by exploring the three worlds and their relationships [21].

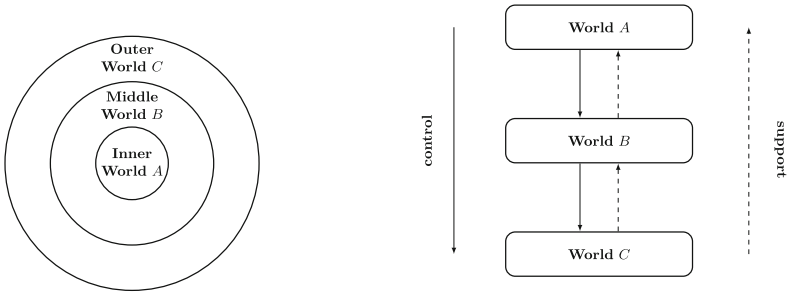
Gu and Zhu [11] proposed a tripartite WSR (wuli-shili-renli) model as a basis of a systems methodology of management. The W (wuli) is about regularities in objective existence, the S (shili) is about ways of seeing and doing, and the R (renli) is about patterns underlying human relations. It is possible to interpret the WSR model based on a triangle of three worlds: W represents the natural world (domains of natural sciences), R represents the human world (human and human society, domains of psychology, social sciences, humanities, etc.), and S represents the applied world (pragmatic problem-solving, human conduct, domains of management science, engineering, operational research, etc.). Theories and knowledge discovered in both W and R worlds are used to guide human conduct in S world, which may change both W and R worlds. To be a better problem-solver, one must integrate the three worlds.

Stern [30] suggested a triadic conception of the reality, in which the reality is conceived and represented as "unified and wholistic as well as differentiated" three worlds: physical world of matter/energy, theoretical world of meaning, and phenomenological world of experience. Furthermore, Stern gave a simplified diagram by enclosing the triangle of the three worlds in a circle representing the unity and wholeness.

One can easily observe both similarities and differences of these three three-world models. Although the contents of the three models are useful and important by themselves, what most interests us is the common triadic structure. On the one hand, the three models have their respective different divisions, understandings, and representations of the reality. On the other hand, they agree upon a three-world triadic structure. It is their agreement on the use of a triangle of three worlds that supports and applies the principles of three-way decision as thinking in threes.

3.4 A Concentric Tricircle Model of Three Worlds

In some situations, we may build three worlds sequentially such that one is on top of another. There are at least two possible ways to depict such a structure [46]. The concentric tricircle of Fig. 3(a) gives us a sense of an inner-outer relationship, or a core-shell relationship, among the three worlds. Typically, an inner world determines an outer world, and the core is more important and serves as a foundation for constructing the outer ones.



(a) Concentric tricircle of three worlds (b) Trilevel of three worlds

Fig. 3. Thinking through three worlds with a concentric tricircle or a trilevel

The inner-outer layered interpretation of a concentric tricircle makes it a commonly used architecture for explanation. For example, in understanding a computer system, the inner kernel represents machine hardware, the middle layer represents system software, and outer layer represents application software. In the Golden Circle leadership model by Sinek [29], the three circles are labeled, respectively, by WHY, HOW, and WHAT. By moving inside-out, a successful leader starts with WHY (i.e., purpose and goals) and moves towards WHAT. Similarly, in the model of behavior change by Clear [6], the three circles correspond to Identify, Processes, and Outcome. We build habits by moving inside-out in the identity-directed way. More examples of three-world thinking based on a concentric tricircle can be found in another paper [46].

Figure 3(b) of three levels gives us a sense of a top-down or a bottom-up relationship among the three worlds. Typically, a world at a higher level controls its lower level and, at the same time, is supported by its lower level. The earlier discussions have shown that three-level models arise naturally in the Venn diagram model and the triangle model of three-world thinking. Trilevel thinking is an important mode of three-way decision. Many examples of trilevel thinking can be found in another paper [44].

4 Three-world Thinking for Building Explanations

The triadic structures of three worlds offer architecture and a scheme for us to make sense of the reality and ourselves. Depending on different contexts and

applications, we may have different interpretations of a triad of three worlds. In this section, I discuss the notion of an SMV (Symbols-Meaning-Value) space [48, 49] as a concrete interpretation of the three-way conception for the purpose of building and communicating explanations.

Weaver [27] insightfully divided communication problems into three categories, which is quoted here:

> Relative to the broad subject of communication, there seem to be problems at three levels. Thus it seems reasonable to ask, serially:
> LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)
> LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
> LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

The three levels focus on different types of problems and answer different types of questions, from easier ones to more difficult ones. In the case of human communication through speaking and writing, we may interpret the three levels by the Words-Meaning-Impact triad. The SMV (Symbols-Meaning-Value) space generalizes Weaver's ideas to a much broader context and provides a structure for trilevel or triadic thinking in many other fields. Considering any theory or model, the SMV space suggests that we need to explain the theory at three levels: the content of the theory, the meaning of the theory, and the utility of the theory[5].

In an attempt to build a conceptual model for explaining data science, I explored a close connection between the SMV space and the widely used DKW (Data-Knowledge-Wisdom) hierarchy [48]. In terms of the three-world thinking, World S is about data (i.e., raw symbols), World M is about knowledge (i.e., meaning of data), and World V is about wisdom (i.e., value from wise use of knowledge). The three-level structure reflects the dependency and transformation between data, knowledge, and wisdom. A conceptual model of data science needs to consider the issues in the three worlds of the data, the knowledge hidden in the data, and the value of the knowledge, as well as the issues arisen from the interactions of the three worlds. Broadly speaking, three goals of data

[5] As an example, we may take a look at the many different interpretations and explanations of a Chinese classic, "I Ching" (The Book of Changes). "I Ching" has shaped every aspects of Chinese ways of seeing, knowing, and living (for example, culture, art, politics, science, etc.) throughout the Chinese history. Many scholars have interpreted and explained, and are continually searching for new interpretations and explanations, this classic text from many different angles. The notion of SMV space may shed a new light by organizing some of the existing interpretations and explanations at the three levels: (1) images and numbers at the S (Symbols) level, (2) meaning and principles at the M (Meaning) level, and (3) living and practice, according to its meaning and principles, at the V (Value) level. Although this organization may not be hundred percent appropriate or accurate, it does provide a good enough approximation in terms of the text itself, the meaning of the text, and the value of the text.

science are (a) to make data a kind of resources through data collection, storage, retrieval, etc., (b) to make data meaningful through data analysis and knowledge discovery, and (c) to make data valuable through practical application of the knowledge in data for making wise decisions and taking the right actions.

In another attempt to build a conceptual model for explaining human-machine co-intelligence, I viewed the SMV space as an architectural system or a metaphorical structure used by an intelligent being to understand and organize itself, its environments, and relationships with others [49]. Human-machine co-intelligence emerges from human-machine symbiosis in the SMV space. There are three fundamental principles of human-machine co-intelligence. The principle of unified oneness: Human-machine co-intelligence is the third intelligence that is based on human intelligence and machine intelligence on the one hand and is above both on the other hand. Human-machine co-intelligence is not possessed by either humans or machines, but through their seamless unification and integration. The principle of division of labor: Human-machine co-intelligence combines the computational power of machines and the cognitive power of humans through proper division of labor. Moving from the World S, to the World M, and to the World V, humans are doing more work and the machines are doing less. The principle of coevolution: Humans and machines mutually adapt to each other, learn from each other, and work with each other as equal partners. Human-machine co-intelligence exploits a mutualism symbiosis in which both humans and machines benefit and, at the same time, avoids a parasitism symbiosis in which one hurts the other. In this respect, in addition to their own SMV spaces, humans and machines share a common SMV space[6]. The notion of SMV space is a structure and a starting point for explaining human-machine co-intelligence.

I now turn my attention to the possibility of applying the SMV space to explainable AI. I have the view that the concept of SMV space suggests a plausible trilevel scheme for constructing an easy-to-understand explanation in explainable artificial intelligence. The SMV triad leads to a trilevel results-meaning-value (RMV) framework of explanation. Like data, the results from a system may be considered as the raw materials that need, or can be used to construct, an explanation. An intelligent system explains its results, outcome, or output (e.g., recommendations, actions, behaviors, etc.), the meaning of the results, and the value of the results at three separate levels. Moreover, at each level, it is possible to apply the ideas of the Venn diagram or the triangle configurations of three worlds to focus on three related questions characterized by the What-Why-How triad. Table 1 summarizes the main features of this 3×3

[6] A few important issues regarding AI and human-machine relations are relevant to the discussion here, such as alignment and control. Christian [5] argued that artificial intelligence systems, in particular machine learning, need to be aligned with human values. Russell [24] pointed out that advances in AI may pose a potential risk to the human race by out of control superhuman AI. Future AI research must ensure that machines remain beneficial to humans and we humans must retain "absolute power over machines that are more powerful than us." By living together in the three worlds of SMV, namely, symbols/data, meaning/knowledge, value/wisdom, humans and machines may coexist in harmony.

architecture of explanations. A 'What' question is about the existence, a 'Why' question is about the reasons/motivations, and a 'How' question is about the processes/applications. By focusing on three fundamental questions of What, Why, and How at each of the three levels, an explanation follows a clearly defined logic, is easy-to-understand, and covers three important aspects.

Table 1. 3×3 architecture of explanations

SMV	Explanation level	Questions
Value	Value	**What** is the value of the results?
		Why are the results valuable?
		How to use the results?
Meaning	Meaning	**What** is the meaning of the results?
		Why are the results meaningful?
		How to interpret the results?
Symbols	Results	**What** are the results?
		Why are certain input/conditions required?
		How does the system derive the results?

A trilevel explanation with three basic questions at each level reflects the principles of triadic thinking. Generally speaking, at a given time, it is possible to focus on the discussion at each level without much interference from the other two levels. In other words, we may need to consider only three questions at a particular level, instead of nine questions at all three levels simultaneously. The labels of the three levels and the three questions at each level in Table 1 may be interpreted more liberally. Depending on different applications, it is possible to use other labels and to ask other types of questions. Nevertheless, the essential components and the structure of the 3×3 architecture remain unchanged. The 3×3 architecture provides a very general framework. In some situations, it may be only necessary to consider some of the nine issues when constructing an explanation. This is particularly true if the results from a system are simple and/or self-explanatory.

5 Conclusion

Three-way decision and three-world conception mutually support each other. On the one hand, three-world models enrich the studies of three-way decision by offering new views, models, and methods. On the other hand, the fundamental philosophy and principles of three-way decision may find new applications in three-world models. In this paper, I explored in brief the connections of three-way decision and three-world conceptions. Thinking through three worlds offers the necessary simplicity and flexibility for building a theory, a model, an argument,

etc. In particular, I examined three models of three-world thinking based on, respectively, a Venn diagram, a triangle, and a concentric tricircle (or a trilevel) organization of three worlds.

I motivated this study by stating that three-way decision is a human-inspired theory. Since humans frequently and naturally think in threes, theories, models, or methods are easy to grasp and understand if they are constructed based on a tripartite architecture. Therefore, explanations from any intelligent systems may be built in a human-friendly way by following a tripartite scheme. It may be fruitful to apply the principles and ideas of three-way decision and three-world thinking to address the issues of the quality and effectiveness of explanations in explainable artificial intelligence (XAI). In this paper, I only presented a proposal for an important research direction, which may be called "three-way decision for explainable AI." Although I gave an outline of a trilevel framework for building explanations based on the notion of an SMV (Symbols-Meaning-Value) space, many fundamental questions remain unanswered. Based on the discussion in the paper, we can explore the new territory of three-way decision and three-world thinking for XAI.

References

1. Achinstein, P.: The Nature of Explanation. Oxford University Press, New York (1983)
2. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access **6**, 52138–52160 (2018)
3. Assagioli, R.: The Balancing and Synthesis of the Opposites. Psychosynthesis Research Foundation, New York (1972)
4. Boer, C.: Thinking in threes: how we human love patterns. Kindle Edition (2014)
5. Christian, B.: The Alignment Problem: Machine Learning and Human Values. W.W. Norton & Company, New York (2020)
6. Clear, J.: Atomic Habits: An Easy & Proven Way to Build Good Habits & Break Bad Ones. Avery, New York (2018)
7. Cowan, N.: The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. **24**, 87–185 (2000)
8. Deakin, M.A.B.: The Name of the Number. ACER Press, Camberwell, Victoria (2007)
9. Dundes, A.: The number three in American culture. In: Dundes, A. (ed.) Every Man His Way: Readings in Cultural Anthropology, pp. 401–424. Prentice-Hall, Englewood Cliffs (1968)
10. Gallo, C.: Talk Like TED: The 9 Public-Speaking Secrets of the World's Top Minds. Sy. Martin's Press, New York (2014)
11. Gu, J.F., Zhu, Z.C.: Knowing Wuli, sensing Shili, caring for Renli: methodology of the WSR approach. Syst. Pract. Action Res. **13**, 11–20 (2000)
12. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z.: XAI - Explainable artificial intelligence. Sci. Robot. **4**, 7120 (2019)
13. Harteveld, C.: Triadic Game Design: Balancing Reality. Meaning and Play. Springer-Verlag, London (2011). https://doi.org/10.1007/978-1-84996-157-8
14. Kaufman, E.L., Lord, M.W., Reese, T.W., Volkmann, J.: The discrimination of visual number. Am. J. Psychol. **62**, 498–525 (1949)

15. Logan, D., King, J., Fischer-Wright, H.: Tribal Leadership: Leveraging Natural Groups to Build a Thriving Organization. Harper Business, New York (2011)
16. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**, 81–97 (1956)
17. Mouton, J.: Understanding Social Research. Van Schaik Publishers, Hatfield, Pretoria (1996)
18. Pawlak, Z.: Rough sets. Int. J. Comput. Inf. Sci. **11**, 341–356 (1982)
19. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
20. Pogliani, L., Klein, D.J., Balaban, A.T.: Does science also prefer a ternary pattern? Int. J. Math. Educ. Sci. Technol. **37**, 379–399 (2006)
21. Popper, K..: In Search of a Better World: Lectures and Essays from Thirty Years. Routledge, New York (1994)
22. Radej, B., Golobič, M.: Complex Society: In the Middle of a Middle World. Vernon Press, Wilmington, Delaware (2021)
23. Ruben, D.H.: Explaining Explanation. Routledge, New York (1990)
24. Russell, S.J.: Human Compatible: Artificial Intelligence and the Problem of Control. Viking, New York (2019)
25. Schneider, M.S.: A Beginner's Guide to Constructing the Universe: The Mathematical Archetypes of Nature, Art, and Science. Harper, New York (1994)
26. Schopenhauer, A.: The World as Will and Idea, Vol. I, II, III, 7th Edition. Kegan Paul, Trench, Trübner & Co., London (1909)
27. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. The University of Illinois Press, Urbana (1949)
28. Shaw, V.N.: Three Worlds of Collective Human Experience: Individual Life, Social Change, and Human Evolution. LNCS (LNAI), Springer, Cham (2019). https://doi.org/10.1007/978-3-319-98195-6_16
29. Sinek, S.: Start With Why: How Great Leaders Inspire Everyone to Take Action. Portfolio/Penguin, New York (2009)
30. Stern, H.W.. In support of a triadic conception of reality. https://harrisstern.medium.com/in-support-of-a-triadic-conception-of-reality-38a784229e9d Accessed 9 June 2022
31. Tall, D.: How Humans Learn to Think Mathematically: Exploring the Three Worlds of Mathematics. Cambridge University Press, New York (2013)
32. Van Melle, W.: MYCIN: a knowledge-based consultation program for infectious disease diagnosis. Int. J. Man-Mach. Stud. **10**, 313–322 (1978)
33. Watson, P.: Ideas: A History, from Fire to Freud. Weidenfeld & Nicolson, London (2005)
34. Webel, C.P.: The World as Idea: A Conceptual History. Routledge, New York (2022)
35. Wei, W.J., Miao, D.Q., Li, Y.X.: A bibliometric profile of research on rough sets. In: IJCRS 2019, LNCS, vol. 11499, pp. 534–548 (2019). https://doi.org/10.1007/978-3-030-22815-6_41
36. Weinberg, S.: To Explain the World: The Discovery of Modern Science. Harper Perennial, New York (2015)
37. Yang, B., Li, J.: Complex network analysis of three-way decision researches. Int. J. Mach. Learn. Cybern. **11**(5), 973–987 (2020). https://doi.org/10.1007/s13042-020-01082-x
38. Yao, J.T.: The impact of rough set conferences. In: IJCRS 2019, LNCS, vol. 11499, pp. 383–394 (2019). https://doi.org/10.1007/978-3-030-22815-6_30

39. Yao, Y.Y.: Three-way decision: an interpretation of rules in rough set theory. In: RSKT 2009, LNCS, vol. 5589, pp. 642–649 (2009). https://doi.org/10.1007/978-3-642-02962-2_81
40. Yao, Y.Y.: Three-way decisions with probabilistic rough sets. Inf. Sci. **180**, 341–353 (2010)
41. Yao, Y.Y.: An outline of a theory of three-way decisions. In: RSCTC 2012, LNCS, vol. 7413, pp. 1–17 (2012). https://doi.org/10.1007/978-3-642-32115-3_1
42. Yao, Y.Y.: Three-way decisions and cognitive computing. Cogn. Comput. **8**, 543–554 (2016)
43. Yao, Y.Y.: Three-way decision and granular computing. Int. J. Approximate Reasoning **103**, 107–123 (2018)
44. Yao, Y.Y.: Tri-level thinking: models of three-way decision. Int. J. Mach. Learn. Cybern. **11**, 947–959 (2020)
45. Yao, Y.Y.: Three-way granular computing, rough sets, and formal concept analysis. Int. J. Approximate Reasoning **116**, 106–125 (2020)
46. Yao, Y.: The geometry of three-way decision. Appl. Intell. **51**(9), 6298–6325 (2021). https://doi.org/10.1007/s10489-020-02142-z
47. Yao, Y.Y.: Set-theoretic models of three-way decision. Granular Comput. **6**, 133–148 (2021)
48. Yao, Y.Y.: Symbols-Meaning-Value (SMV) space as a basis for a conceptual model of data science. Int. J. Approximate Reasoning **144**, 113–128 (2022)
49. Yao, Y.Y.: Human-machine co-intelligence through symbiosis in the SMV space. Appl. Intell. 1–21 (2022). https://doi.org/10.1007/s10489-022-03574-5