Arbitrarily applicable relational responding as non-axiomatic logical reasoning

Robert Johansson[1,2], Jonas Ramnerö[2], & Arne Jönsson[1]

[1]*Department of Computer and Information Science, Linköping University, Sweden*

[2]*Department of Psychology, Stockholm University, Sweden*

robert.johansson@liu.se, jramn@psychology.su.se, arne.jonsson@liu.se

Stimulus equivalence

In the late 1960's, Murray Sidman was working with language comprehension with severely developmentally disabled individuals. Unexpectedly, he discovered that if subjects were successfully taught to match pictures and printed words to dictated words, and to name pictures, they would without explicit training learn how to match printed words to pictures, match pictures to printed words and to "read" (i.e., name words). From a behavioral psychology point of view, this was very interesting, as it demonstrated a clear example of emitted behavior without a history of reinforcement. This discovery has resulted in over 40 years of research on *stimulus equivalence* (Sidman, 2009). Stimulus equivalence is a behavioral phenomenon that has only been observed in humans with verbal abilities (with one possible exception, of a california sea lion "Rio"). However, generally no non-human animals seem to have been able to do this (Zettle et al., 2016). The typical way to study stimulus equivalence is with the help of matching-to-sample experiments. In such experiments, participants are exposed to series of arbitrary stimuli (e.g., nonsense symbols) where the task is to match a certain symbol to a given sample stimuli. Such experiment is an example of *relational responding*. That is, the task for a participant is not to emit a response in relation to a certain stimulus. It is rather to respond to the relation between symbols.

A formal definition of stimulus equivalence follows. Assume three nonsense symbols, which we for simplicity will refer to as A, B and C (they might be nonsense words, pictures, or something else). Within a given experiment (like the matching-to-sample), participants are taught to select B rather than some other option in the presence of a sample A (i.e., the relation A → B will be established). In the same way C is trained as the correct response in the presence of B (B → C). After these relations have been trained, without training in other relations, participants demonstrate an increased probability of selecting A from a set of options when B is presented as a sample (B → A; *symmetry*), selecting C when A is displayed (A → C; *transitivity*), selecting A when C is displayed (C → A; *equivalence*), and also the trivial case of selecting A when A is displayed (A → A; *reflexivity*).

Demonstrating symmetry and equivalence are examples of *derived relational responding*, as these relations are not directly taught but instead derived. Prior to the research by Sidman and colleagues the emergence of these derived stimulus relations was not expected in similar experimental setups. The stimulus equivalence phenomenon opened up for a new way of studying symbolic relations (i.e., how a word "represents" an object in language), and supported the idea that derived stimulus relations were an important component in language and cognition. Importantly though, the idea is not new. The abstract concept of sameness or equivalence has long been regarded as *"the very keel and backbone of our thinking"* (William James, *Principles of Psychology*, 1890/1998, p. 459).

Arbitrarily applicable relational responding

In the late 1980's, the developers of *Relational Frame Theory* (RFT; Hayes et al., 2001) started to ask questions on what was beyond equivalence, for example: What kind of derived relational responding based on other relations than equivalence are human beings capable of? And if so, would such responding also be reflexive, symmetrical, and transitive?

Consider the following statement: *"A is more than B and B is more than C"*. Not only are the AB and BC relations specified, we immediately derive the BA, CB, AC, and CA relations. Hence, we are able to answer a question such as *"Is C more than A"?* (The answer would be *"No"*) Consider an elaborated version of this example. Imagine someone standing in a coffee shop looking at the menu with words such as "Espresso", "Americano" and "Caffé Au Lait". If the person looking at the menu has no experience of these brands, he/she might ask "How are these related regarding strength?". An answer might be "the Espresso is stronger than Americano, and the Americano is stronger than the Caffé Au Lait". The person asking will immediately be able to derive the other relations, for example that Caffé Au Lait is less strong than Espresso. Furthermore, let's say that the person tastes Americano, we could in behavioral terms say that Americano acquires various *stimulus functions*, such as taste and smell. Importantly though, what will happen is that the stimulus functions throughout the whole Espresso-Americano-Caffé Au Lait-network will be transformed due to the invormation given. For example, the person might be able to imagine the strength of taste on Espresso, despite no actual experience. Such "derived experience" could affect future decision making for the customer in the coffee shop.

How is the above related to stimulus equivalence? Clearly, both phenomena is about derived stimulus relations. For stimulus equivalence, reflexivity (A is the same as A) was required. This seems not to be the case for more than/less than (A is not more than A). Regarding symmetry (if A=B then B=A), there seem to be modified versions of this for more than/less than relations (If A>B then B<A). Transitivity seems to behave similarly (if A>B and B>C then A>C). Consider another relation such as opposition, and someone learns *"A is the opposite to B, and B is the opposite to C"*. This relation is not transitive, as someone will be able to derive that A is actually the same as C. In RFT, these "generalized versions" of symmetry and transitivity are labeled *mutual entailment* and *combinatorial entailment*, respectively (if A>B>C then the relation B<A is mutually entailed, and C<A is combinatorially entailed).

Several broad classes of relating have been discovered: *Coordination* (equivalence, sameness), *Opposition*, *Distinction* ("is different from"), *Comparison* (e.g., more/less or bigger/smaller), *Hierarchy* (contains/member of), *Temporal* (before/after), *Spatial* (Here/There), and *Deictical* (relations in terms of the perspective of the speaker; a combination of interpersonal relations like I/You with spatial and temporal) (Zettle et al., 2016). Expressed in RFT terms, all of these overall patterns of relating have the properties of *mutual entailment*, *combinatorial entailment* and *transformation of stimulus function*. In RFT, these patterns with these properties are referred to as *relational frames*.

Now consider this scenario: Someone new to a country learns that three never before seen coins A, B, and C are ordered along a comparative dimension such as *"A is worth more than B that is worth more than C"*. Having learned this enables the person to operate efficiently in several potential decision making scenarios that involves money in the new country. What's crucial about this example is the fact that the comparative relations between A, B and C are along an arbitrary dimension of worth, rather than for example size, weight or other directly observable properties. This is an example of when a relational frame of comparison is *arbitrarily applied*. In RFT terms, it is the *contextual cue* of "is worth more" that controls the application of the frame of comparison. Had someone instead said "A is equal to B", we could have expected the frame of coordination to be applied. This illustrates the arbitrary nature of relational frames.

In summary, *arbitrarily applicable relational responding (AARR)* is defined as abstract response patterns, that has the properties of mutual entailment, combinatorial entailment and transformation of stimulus functions, and that are controlled by contextual cues. From an RFT perspective, cognition is not a mental event that mediates between environment and behavior. It is rather a behavioral event (AARR), and hence, it can be studied and understood within a behavioral psychology framework, using experiments such as the matching-to-sample task. Another way to put it: arbitrarily applicable relational responses are what "minds" are full of, and when we speak of "cognitive" phenomena (such as "thinking", "planning", "remembering", "decision making") we are referring to complex instances of relational framing that are more or less evident under different environmental conditions (Zettle et al., 2016).

NARS and non-axiomatic logic

It is clear that AARR is a domain-independent process, potentially occuring at many levels, and seems to be involved in many "cognitive" functions. One example of a system that aims to reason about such processes is the general-purpose intelligent system *NARS* (Non-Axiomatic Reasoning System). NARS is designed to be adaptive and to work with insufficient knowledge and resources. Its various cognitive functions are uniformly carried out by a central reasoning-learning process following a "non-axiomatic" logic (Wang, 2013).

NARS makes use of a formal language, "Narsese", for its knowledge representation, and this language is defined using a formal grammar (Wang, 2013). The system's logic is developed from a so-called *term logic*. Statements in this logic have the form subject-copula-predicate. The smallest element that can be used as one of these components is referred to as a *term*. In Narsese, the most basic statement is the *inheritance* statement, with the format *"S → P"*, where S is the subject term, and P is the predicate term. The "→" is the inheritance copula, which is a reflexive and transitive relation. The intuitive meaning of *"S → P*" is *"S is a special case of P"* and *"P is a general case of S"*. For example, the statement *"bird → animal"* intuitively means *"Bird is a type of animal"*. Importantly, such a statement doesn't say anything about the meaning of the terms itself – it merely states the relationship. Terms can be grouped together in various forms of sets, for example *{Cat, Dog, Giraffe}*, enumerating instances, or *[yellow, tall, four_legged]*, enumerating attributes. Various arbitrary relations can be represented, for example the relation *"Silvia is the mother of Victoria"* is in Narsese represented as *"({Silvia} × {Victoria}) → mother-of"*. A statement such as *"Tim knows snow is white"* can be represented as a higher-order statement *"{Tim} → (know / ⋄ {snow → [white]})"*, where the statement *"snow → [white]"* is used as a term. Beside the inheritance copula ('→', *"is a type of"*), Narsese also includes three other basic copulas: *similarity* ('↔', *"is similar to"*), *implication* ('⇒', *"if-then"*), and *equivalence* ('⇔', *"if-and-only-if"*). The last two copulas are "higher order", meant to be applied to statements themselves.

Furthermore, NARS can reason on *events*, that are described as statements with temporal attributes. For example, *"event E1 happens before E2"* is described in NARS as *"E1 /⇒ E2"*. Finally, *procedural operations* in the system are events realized by the system itself. These operations are typically executable commands or procedures of the system. Formally, an operation is an application of an operator on a list of arguments, written as $op(a_1, \ldots, a_n)$ where *op* is the operator, and $a_1, \ldots, a_n$ is a list of arguments. Such an operation is interpreted logically as statement *"(× {SELF} {a_1} . . . {a_n}) → op"*, where *SELF* is a special term indicating the system itself, and op is a term that has a procedural interpretation. For instance, if we want to describe an event *"The system is holding green brick nr 2"*, the statement can be expressed as *"(× {SELF } {green_brick_2}) → hold"*.

Goal of this work

The primary aim of this work is to investigate if NARS can do AARR with gradually increasing complexity, and under which conditions this is made possible. During the presentation, we will describe a research plan, starting with stimulus equivalence, and then continuing with more advanced relational responding. How this work potentially could be beneficial for RFT research and for research on intelligent systems will be discussed, as well as potential future applications.

References

Hayes, S.C., Barnes-Holmes, D. & Roche, B. (ed.) (2001). *Relational frame theory: a post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum Publishers.

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *The Analysis of verbal behavior*, *25*, 5-17.

Wang, P. (2013). *Non-axiomatic logic: a model of intelligent reasoning*. New Jersey: World Scientific.

Zettle, R.D., Hayes, S.C. & Barnes-Holmes, D. (ed.) (2016). *The Wiley Handbook of Contextual Behavioral Science*. Chichester, West Sussex, UK: Wiley.