

## **The Structure and Interpretation of Realms of Thought<sup>1</sup>**

### **I. Introduction**

The complex architecture of the mind developed by Marvin Minsky in his forthcoming book The Emotion Machine articulates many simple yet vital structures that when used in concert can lead to the complexities of the mind. One set of very important structures are realms of thought, which are sets of related ways to interpret and mentally interact with things. For example, I might think of a bicycle in terms of being a machine, being a conveyance, being a possessable object, being beautiful, and so on. Each of these different views or opinions associated with the bicycle are representative of different realms of thought. Respectively, drawing from and adding to names Minsky gives to some realms of thought, they are the physical, transportation, dominion, and artistic realms of thought. Realms of thought provide a basic way of interpreting various parts of the world, and are associated with certain mental processes related to that realm of thought.

Minsky discusses realms of thought in terms of their utility in problem solving. Often when attempting to solve a problem or attain a goal, the mind will have to adapt its problem solving strategy. This adaptation may require the mind to quickly activate and effectively use mental resources which were previously dormant to solve a problem. Realms of thought provide for these shifts. For example, if I want to go to the store, and I have a bicycle, I will think about how to get to the store and use my transportation realm of thought to conclude I will ride my bicycle to the store. If, however, I realize my bicycle is not working, say, because the chain fell off the sprocket, I need to switch to the physical realm of thought and try to fix it. Devoting mental resources to analyzing various methods of travel to the store will not fix my bicycle; I

---

<sup>1</sup> A reference to the great work of Gerry Sussman, specifically Structure and Interpretation of Computer Programs, MIT Press, 1996

should quickly stop thinking about transportation and start thinking about how I can fix my bicycle.

Shifting mental resources can be described without using structures like realms of thought. However, categorizing sets of related mental processes into realms of thought enables clear and useful discussion of how mental resources are allocated. As such, I will initially assume they exist as actual mental structures, and will return to the issue of whether realms of thought are actual structures or merely useful artificial abstractions in Section IV.

Herein I present a detailed look at realms of thought, expanding on their outline provided by Minsky. To do so, I define a realm of thought to be a set of related mental processes which use a common representation to provide some understanding of a situation. A realm of thought's primary function, therefore, is to activate certain mental processes and claim mental resources for those processes, and not to actually interpret a situation itself. This means that other mental processes, such as goal determination, credit assignment, and memory, exist separately of realms of thought, though they may interact with and be affected by particular realms of thought. Generally, only one realm of thought is primary, meaning highly active, at a time, although other realms of thought may be using a few mental resources in the background.

I will first present a brief survey of related work to ground my arguments and facilitate analysis. Next, through a sample scenario, I will relate realms of thought to other mental processes not discussed in the context of realms of thought by Minsky. I will then explore new ideas about realms of thought, including if and how they might change or be created, whether or not realms of thought physically exist, and whether or not it is important that they actually exist.

## **II. Related Work**

Not much research has been done directly on realms of thought, with much of the discussion being confined to Marvin Minsky and his research partners, due in part to the newness of the idea and in part to the radical departure from traditional artificial intelligence his work represents (Minsky 1986, Minsky forthcoming). However, Push Singh, one of Minsky's graduate students, developed a system embodying many of the principles outlined by Minsky, including realms of thought, as well as contributing significant work to artificial intelligence in his own right (Singh

2005). Also important to realms of thought is Minsky's concept of frames, which allow realms of thought to represent the same objects in multiple ways (Minsky 1974).

More work has been done on multiagent systems, a class of systems which includes systems similar to Minsky's design (Decker & Lesser 1995, Grosz 1999, Stroulia & Goel, 1995). These systems demonstrate that simple parts can be used to make complicated machinery. Erik Mueller has specified a story understanding system which incorporates some of Minsky's ideas, including realms of thought (Mueller 1999). Finally, recent research suggests that much of human thought is domain specific, meaning that a psychological basis for examining realms of thought exists (Hirschfeld & Gelman, 1994). However, as previously mentioned, much of this work, with the exception of Minsky's and Singh's work, does not provide insight into realms of thought themselves, but rather a justification for investigating them.

### **III. Realms of Thought and the Rest of the Mind**

Realms of thought, though significant, are by no means the primary or broadest structure in the mind. Other processes and structures perform other very important tasks essential to human-level thought, and realms of thought interact with them in various ways. I will hypothesize how some of these interactions might occur through a hypothetical scenario.

#### **A. *Joan Movie to a New Town***

Joan is twelve years old, and her father was just transferred from Los Angeles to New York. Joan has never been to New York, and has no friends there. At first she is nervous and a little sad about leaving the only place and people she's ever known, but after talking to her parents, she soon becomes excited at the chance to meet new people and reinvent herself. As she prepares for the move, she thinks about how she wants to act in New York. She knows she wants to make new friends, but what kind of friends will she have? Does she want to be the artist? The athlete? The smart kid? The same as she was in California? Eventually she decides: she'll be essentially the same, except that now she'll also become a swimmer, and she'll be a little more confident, since no one in New York will know about that embarrassing moment in second grade. Soon Joan is ready to move.

Two months later Joan and her family move to New York, and Joan sets about her plan to make friends and become a new person. She begins meeting children on her block and in her school, and even meets a boy she likes. She also tries out for the swim team, and, after the first few practices, she has improved in her swimming well enough to make the team, as well as a few more friends. And because of her new confidence, and her favorite boy's membership, she also decides to join the debate team. Within a few months, she has made many friends and feels satisfied with her new life. However, when her best friend begins dating the boy she likes, Joan is crushed. She feels betrayed and hurt. It is only when the boy breaks up with Joan's friend, and the friend turns to Joan for consolation, that Joan realizes she has made a true friend. It wasn't the swimming or the erased embarrassing moment which made a real friend; it was just Joan being there for a hurt friend, despite that friend's earlier transgressions. Joan decides in the future this is how she will make friends: be there for people when they need you.

## ***B. What did Joan Learn***

This story provides fertile ground for analysis in the same fashion as the analysis of various stories in The Emotion Machine, and as such, that analysis will not be presented here. Rather, I will focus on analyzing how the realms of thought Joan is using at various points in the story interact with her other mental faculties.

Over the course of the story, Joan changes her mind about an idea, specifically how she should act to make friends. The process of changing her mind involves several realms of thought through the story, as well as many other mental processes. Joan's realms of thought affect other her other mental processes, something not fully explored in Minsky's book. We will examine how Joan's realms of thought may affect her mental processes, specifically, her processes of goal acquisition and credit assignment. These processes interact with both each other and Joan's realms of thought.

### **1. Goal Acquisition**

Throughout the story, Joan is pursuing one main goal, to make new friends. To that effect, she creates subgoals designed to move her closer to that goal. For example, she decides to join the swim team, and to not tell people about an embarrassing moment before she even departs for

New York. But how does she create these subgoals?

Joan's mind contains a set of critics, which look for discrepancies between Joan's desires and the real world, and selectors, which propose solutions to the discrepancies noticed by the critics. These critics and selectors use narratives Joan has experienced to facilitate the formation of plans based on making analogies to similar experiences Joan has had in the past. By recursive use of critics and selectors, Joan can create reasonable long-term plans and goals. This description is the basic process of goal acquisition and planning outlined in The Emotion Machine.

The critics, selectors, and stories, however, can be grouped in a manner similar to realms of thought. While many critics and selectors are general purpose, some may be domain specific, and most, if not all, stories are relevant only to certain domains. For example, Joan's decision to join the swim team may have gone something like this:

- (1) Joan wants to make friends (critic)
- (2) Participating in an extracurricular activity can lead to making friends (selector referring to a story about making friends in an activity)
- (3) Joan needs an activity to do (critic referring to the fact that Joan is not in an activity)
- (4) Joan knows how to swim (selector referring to a story about Joan swimming)
- (5) How does swimming allow Joan to participate in an activity (critic referring to a story about swimming independently of other people)
- (6) Joan can join the swim team (selector referring to a story about swim teams being group activities)

The stories used in this example contain information useful in some domains and not useful in other domains. For example, the story used by the selector in step (2) conveys the information that Joan can make friends in extracurricular activities. This story contains information about social dynamics, and can therefore be associated with the social domain. The story in step (4), however, conveys the information that Joan can swim, which is relevant to her physical abilities, and thereby the physical domain, but in and of itself, this story includes no information about Joan's social abilities. Thus, over the course of this example, the domain of the information conveyed by the stories used changes, and over the whole story is

- (1) Social
- (2) Social
- (3) Social, Physical, and Mental
- (4) Physical
- (5) Social, Physical
- (6) Social

The nature of information carried by the stories used at each step is related to two things: the nature of the overall goal, in this case a social goal, and the content of the information in the last story used in the process. For example, step (2) suggests that Joan join an extracurricular activity. This suggestion poses the question of whether or not Joan is involved with extracurricular activities, a social and possibly physical and/or mental issue. Step (3) responds with information related to social, physical, and mental issues, and poses the question of what Joan can do to join an activity, related to physical and/or mental issues. Step (4) responds with information about physical issues, and poses another question, and so on. Thus, the appropriate type of story Joan needs can be guessed based on the nature and content of the very last issue Joan addressed.

Due to evolutionary pressures, Joan should be able to recall stories to make plans quickly to increase her chance of survival. Anecdotally, we can observe that people can create most goals and plans in familiar situations quickly, sometimes almost instantly, while creating goals and plans in unfamiliar or complicated situations can take longer. So, for efficiency, Joan's mind should be able to quickly change the set of stories she is examining in a particular situation based on the nature and content of the issue she has just addressed. Realms of thought provide this functionality.

Recall that our definition of realms of thought states that they activate a set of related mental processes and acquire resources for those processes. Among the resources controlled by realms of thought are the narratives Joan uses to create goals and plans. When, in step (3) above, Joan's critic presents a physical problem, Joan's mind switches to the physical realm of thought, preparing the stories related to Joan's physical abilities for use before the selector of step (4) actually requests them. This process is similar to the process pre-fetching the contents of a familiar room prior to actually entering it as described by Marvin Minsky. The pre-fetching of

stories allows Joan to quickly make decisions when she has an appropriate story for the situation; that is, when the situation is familiar. However, if Joan does not have an appropriate story in the realm of thought she expected she would need, meaning the situation is unfamiliar to her, she will have to look elsewhere for something she can use, making her goal acquisition process take longer. This fact explains why people can make goals much faster in familiar situations as opposed to unfamiliar ones. Realms of thought interact with goal acquisition through aiding the speed and efficiency with which the goal acquisition process can operate.

## **2. Credit Assignment**

Once Joan has accomplished her goals, she must determine what actions and thoughts allowed that accomplishment. Over the course of the story, Joan at first believes she has accomplished her goal, and then later realizes she hasn't, only to finally end up truly accomplishing her goal. In both cases where Joan believes she has accomplished her goal, credit assignment occurs so that Joan can act in the future in ways which were successful in the past. The actions and thoughts she selects as responsible for her success are those which occurred within the realm of thought corresponding to the domain of her goal.

The first time Joan believes she has accomplished her goal, when she has made friends after a few months in New York, she must select the actions and thoughts which were significant to her attaining her goal. Her goal was a social one, to make friends. After she has done so, she credits her thoughts in the social realm of thought with the success. For example, one action Joan took which led to her making friends was the decision to join the swim team. As seen in the previous discussion, this decision was composed of several smaller decisions relating to the social, physical, and mental realms of thought, and so it is possible any of these smaller decisions were the key to her making friends, without which she would have failed. However, Joan gains the benefit of future use of this story for reasoning by analogy if she can generalize it rather than credit each and every decision she made. So, because her goal was a social one, she credits her social domain thoughts and actions with helping her attain her goal. She credits the social act of joining the team with helping her make friends, not the physical act of swimming, which could lead her to believe in the future that the act of swimming itself will help her make friends.

Joan can make her credit assignment process simpler and more powerful than examining every decision she made, determining its domain, and deciding whether or not to credit that decision with helping her to her goal. Instead, she can use realms of thought to aid her credit assignment. Joan can credit all decisions which were acted upon without causing her to reexamine the problem made within the realm of thought with solving the problem, and thus correctly credit a large portion of the causal decisions used. Her credit assignment can be understood by analogy. Pretend that any time Joan takes an action about something within the domain of her problem, she writes it down. If that action causes Joan to return to the social realm of thought to deliberate the same problem, meaning her action failed, she erases that action. At the end of the process, Joan has a list of the major decisions and actions she took to solve the problem.

In the story, Joan had a social problem, and will credit the social domain decisions she made with the solving of that problem. She will credit the actions within the social domain which did not require her to reexamine a particular problem with the success. She credits meeting people on her block and at school, not telling people about her embarrassing moment, and joining the swim team and debate teams with her success. She now believes that meeting people, keeping embarrassing things secret, and joining teams can help her make friends, a fairly accurate analysis at this point in the story. Through restricting herself to crediting only those actions brought about by decisions within the same realm of thought as the domain of her problem, Joan can quickly, easily, and accurately credit her success. However, this process is not foolproof, as we shall soon see.

Later in the story, Joan realizes she was wrong about her first assessment of the status of her goal, and only later truly completed her goal. Because she realizes she did not succeed earlier, she must now reevaluate her credit assignment. Joan had earlier credited her initial social actions, like hiding an embarrassing moment and joining the swim team, with her success. Now, she must reevaluate her credit assignment because she has been forced to reexamine the same problem again. She now credits being there for a person in need, a social domain action, with helping her accomplish her goal. While the actions she took previously certainly aided her in making friends, they were not the actions responsible for her success in the same way that knowing how to swim, while an important part of her success, was not responsible for her making friends. She might

realize that her earlier credit assignment is more applicable to the goal of meeting people and making acquaintances, but not to the goal of making friends. The important action involved in making friends, the action deserving credit, is being there for a person in need. She now has a better understanding of how to make friends, and can use this experience in the future.

An issue brought up by her new credit assignment is that being there for someone is a complicated concept. It is not as simple of a social action as joining the swim team; it involves the emotional and mental realms of thought as well. However, these other realms of thought are used in the act of being there for someone, while the act itself is a social domain act. These other realms of thought are used in the act of being there for someone in the same way that the physical realm is involved in Joan's social decision to join the swim team; the act of being there for someone is a social act depending on other domains for success. Being there for someone is itself a goal, with a credit assignment structure. Because Joan credits the social act of being there for someone with her success, she is indirectly crediting parts of her emotional and mental realms of thought with making friends. This fact allows complex interactions among the realms of thought through credit assignment. Joan may have experiences and thoughts within the emotional and mental realms in the course of making friends through her actions of being there for someone, explaining some of the reason why people often feel emotional ties to their social contacts. Realms of thought affect credit assignment by filtering actions based on the domain in which they are relevant, and through credit assignment structures, realms of thought can interact with each other in some circumstances.

#### **IV. Realm of Thought Dynamics**

Realms of thought must be created and develop at some point over the life of a human; there is no homunculus within the zygote full of premade realms of thought. But how do they arise, and at what point, if ever, in the life of a human do they become static?

##### **A. *Development of Realms of Thought***

The mutability of realms of thought cannot be assumed. Realms of thought activate a set of related mental processes and secure mental resources associated with those processes. Thus, a

change to a realm of thought technically appears to be a change in the manner in which it activates processes and accesses resources. While this definition is technically the true definition of a change to a realm of thought, it is not that useful. The action of activating resources and processes is most likely a simple one, and changing it does not seem to be something which would occur often, if ever. Thus, we will define a change to a realm of thought slightly differently.

Although changes to the actual realm of thought itself may seldom occur, changes to the content of the realm of thought occur frequently. People learn all the time, and that learning affects the mental processes and resources associated with a realm of thought. Learning how to throw a football affects what processes (critics and selectors) and mental resources (narratives) can be used by the physical realm of thought. Learning how to integrate a function affects the processes and resources of the mental realm of thought. I will define a change to a realm of thought to be a change to the set of processes and resources available to a realm of thought.

In this light, realms of thought clearly and often change. Learning itself is embodied in changes to realms of thought, because the ability to observe and store information for later manipulation and purposeful use is the definition of learning. Realms of thought provide this functionality through categorizing information according to different domains so that it can be later used to make plans, decisions, goals, and so on. See Section III for an example of learned information (in the form of narratives) being used to form goals.

### ***B. Creation of Realms of Thought***

So realms of thought change, but can they be created? Certainly at some point they are created; again, there is no homunculus in the zygote. But whether people are born with a fixed set of realms of thought or are capable of creating them is a tougher question.

We first conduct two thought experiments:

- (1) Suppose a child was born into an atheist family and has no experience with religion until entering school at age 6. He is then exposed to religion, and eventually becomes very religious and joins the priesthood.

- (2) Suppose another child, this time a teenager, enters school and begins learning Latin, a foreign language he has never been exposed to. Within a few years, he becomes fluent, and even sometimes thinks and dreams in that language.

Both of these thought experiments drive at the same point, that people can learn new categories of information that they were not born with or exposed to while very young. Most people can recall learning entire paradigms and ways of thought at many points in life, for example learning how to think like a computer scientist. Over the course of learning these things, people acquire vast amounts of new processes (critics and selectors) and resources (narratives). But does this learning represent the creation of a new realm of thought? We can view our experiments in two ways:

- (1) The child developed a new realm of thought, the religious realm of thought

-OR-

The child added a large number of mental processes and resources to some preexisting realms of thought, such as the mental and emotional realms.

- (2) The child developed a new realm of thought, the Latin realm of thought

-OR-

The child added a large number of mental processes and resources to some preexisting realms of thought, such as the mental and communication realms.

In the first explanations, we claim that eventually a new realm of thought was created. Sufficient amounts of new mental processes and resources were added to the child's mind that a new realm of thought has spawned from them. The creation of the new realm of thought need not necessarily be a decision made by some part of the child's mind; it could arise simply by the fact that when the child thinks about ,say, religion, a certain set of his mental processes and resources, those concerning religion, are frequently activated in his mind. Over time, these processes and resources could become associated together to the point that they become a new realm of thought.

In the second explanations, we argue that the child has only modified existing realms of thought, although the modifications may be significant. Some realms may be very different from the way they once were, but no new realms of thought, that is, no new structures which activate sets of

mental processes and resources, have been created. One may call for convenience a subset of the realms of thought a subrealm of thought, as in the Latin subrealm of thought, but this does not involve the creation of a realm of thought. It may follow from this argument that at a certain point in one's life, the set of realms of thought become fixed. Realms of thought may even be a set of genetically specified structures, making creation of new ones impossible in a single organism. Biological evidence suggests that regions of the brain are devoted to specific tasks, and perhaps this is a justification for this argument.

Deciding which view is correct, or if another view is correct, is a difficult task based on the evidence on hand. More experiments, real ones, need to be run, and deep contemplation about the nature of learning and realms of thought must be conducted. Thankfully, the answer of this task does not matter.

### ***C. Why the Answer Does not Matter***

Either view of the child is correct. Realms of thought activate related sets of mental processes and resources. That is all they are, that is all they do. Because both explanations of what happened to the child allow that he learned something, that he added new mental processes and resources, they are equivalent. The fact that one explanation calls what happened a new realm of thought and the other does not is irrelevant.

The immediate argument is that realms of thought have been assigned tasks and duties in this paper, and have been argued to provide benefits, evolutionary and otherwise, and generally argued to exist. This is a valid argument. But if we now move beyond discussion of what realms of thought do and why we have them to actually examine what may be behind realms of thought, we find their actual structure and existence is unimportant, and refute this argument.

Remember again we have defined a realm of thought as a structure which activates a set of related mental processes and mental resources, but does not actually determine when it is applicable to a situation in and of itself. In that sense, a realm of thought acts like a lieutenant, activating its troops, its processes and resources, when instructed to do so. One might imagine that some classifier or set of classifiers determines in what domain a particular problem or question exists, and then activates a realm of thought of the corresponding domain. There are

many ways a realm of thought might theoretically activate resources, including opening certain neural pathways, preparing appropriate knowledge for loading into memory, routing information to processes and resources, or releasing chemicals which activate certain neural circuits. From a cognitive science standpoint, determining if realms of thought a physical structures and how they work would be interesting.

But realms of thought as structures do seem complicated. The constant changes in processes and especially resources within a realm of thought would require realms of thought to be very flexible, and the vast number of connections made by a realm of thought may make reliability an issue. This is not a refutation of realms of thought in itself, but does point out that they would be complex if they do exist. In general, evolution has favored the simplest (i.e., most energy efficient) structures. This top-down design for realms of thought requires that they be capable of continual change, either in what neural connections they make, or what cells are affected by signals or chemicals they release. Maintaining the stability and accuracy of such networks would require a significant amount of energy, especially for large realms of thought like the physical realm of thought. If there were a simpler, bottom-up approach in which processes and resources “knew” in what domain they were, a better design might be achieved.

A system called Pandemonium was a system which did just that (Selfridge 1988). Various agents, called demons by Selfridge, processed data based on a system in which the agents were aroused to various degrees by the data. Those agents who were most aroused by particular data were used to interpret that data. I propose a similar system may be at work here. Each narrative, the major mental resources involved with realms of thought, is experienced in a given context related to a goal or perceived goal. That goal falls within a certain domain. If, when a narrative is acquired, the domain in which it was observed is encoded with it, then to later find all narratives within a certain domain, all that needs to be done is for the brain to determine which narratives are aroused by the appropriate domain. The narratives themselves control when they are activated, creating the equivalent of a realm of thought. A similar idea can be applied to critics and selectors, the mental processes associated with realms of thought. Thus, functionality of realms of thought are fully handled through a simple bottom-up mechanism. Further, this design adds fail-soft capability through the fact that if one resource breaks, the rest of the system is

unaffected. In contrast, in a top-down design, if the lieutenant breaks, the entire realm of thought is affected.

Going back to the original question of this section, which explanation of what happened to the child is correct, using the bottom-up model, we now see the answer is irrelevant. The difference between the two explanations is not one of substance but of definition. Realms of thought are an abstraction, a useful way for talking about a behavior. But, like other abstractions, they can be dangerous because they can obscure key details, and also become viewed as not an abstraction of the truth but the truth itself.

It is certainly possible that I am wrong, and realms of thought exist physically in the brain in a top-down structure. Experimentation is needed to verify these claims in a cognitive science context. However, from a computer science perspective, realms of thought have been shown here to be implementable in a bottom-up fashion, leaving them not as structures but design ideas. If the goal of artificial intelligence is to create something equivalent to human intelligence, but not necessarily imitating it, this design is more than sufficient.

## **V. Conclusion**

We have examined in this paper many aspects of realms of thought, from Minsky's original outline of what they might be to exploring if they might merely be a useful abstraction. Key points about realms of thought include their interaction with other mental processes, such as goal acquisition and credit assignment as illustrated in a story, the mutability of realms of thought, if realms of thought can be created, whether or not they actually do exist, and whether or not their actual existence matters. More elaboration and exploration of what realms of thought are and what they can do is certainly warranted; like many of the ideas in [The Emotion Machine](#), there are too many ideas and not enough people thinking about them. Other further research might include implementing a few different strategies for constructing realms of thought and comparing them, as described in the last section. Much work needs to be done, but the end result of this and similar research looks promising.

## VI. Works Cited

- Decker, K. S. and Lesser, V. R. (1995). Designing a Family of Coordination Algorithms. *Proceedings of the First International Conference on Multi-Agent Systems*.
- Grosz, B. J., Hunsberger, L., and Kraus, S. (1999). Planning and Acting Together. *AI Magazine*, 20(4), 23-34.
- Hirschfeld, Lawrence A. and Susan A. Gelman, eds., (1994). *Mapping the mind: Domain specificity in cognition and culture*  
New York, NY: Cambridge University Press.
- Minsky, M. (1986). *The Society of Mind*. New York: Simon and Schuster.
- Minsky, M. forthcoming. *The Emotion Machine*.
- Minsky, M. (1974). A Framework for Representing Knowledge. *Artificial Intelligence Memo 306*, <https://dspace.mit.edu/bitstream/1721.1/6089/2/AIM-306.pdf>
- Mueller, E. T. (1999). *Prospects for in-depth story understanding by computer*. CogPrints cog00000554.
- Singh, P. (2005). EM-ONE: An Architecture for Reflective Commonsense Thinking. Ph. D Thesis
- Selfridge, O. G. (1988). Pandemonium: a paradigm for learning. In *Neurocomputing: Foundations of Research*, J. A. Anderson and E. Rosenfeld, Eds. MIT Press, Cambridge, MA, 115-122.
- Stroulia, E., & and Goel, A. (1995). Functional Representation and Reasoning in Reflective Systems. *Journal of Applied Intelligence*, 9(1), 101-124. Special Issue on Functional Reasoning.