



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Representational Dynamics in Problem Solving“

verfasst von / submitted by

Benjamin Angerer

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (M Sc.)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 013

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Joint Degree Programme MEi:CogSci Cognitive Science

Betreut von / Supervisor:

ao. Univ.-Prof. Dipl.-Ing. Dr. Franz-Markus Peschl

Benjamin Angerer, B.Sc.
CURRICULUM VITAE

PERSONAL DATA

Name	Benjamin Angerer
Nationality	German
Birth date and place	25th July 1985 in Buchloe, Germany
E-Mail	bangerer@uos.de

RESEARCH INTERESTS

Complex Problem Solving – Analogy-Making – Representational Change – Mathematical Cognition – Constructivist Epistemology – Qualitative methodology – Inductive Programming – History of Psychology – History of Computing

EDUCATION

2013 – 2014	Erasmus exchange year, University of Ljubljana (project work on phenomenology and protocol analysis with Prof. Urban Kordeš)
2012 – 2017	Studies in Cognitive Science (M.Sc.) Thesis title: “Representational Dynamics in Problem Solving” (supervisor: Ao. Univ.-Prof. Dr. Markus Peschl). MEi:CogSci Master's Programme, University of Vienna.
2007 – 2012	Studies in Cognitive Science (B.Sc.). Thesis title: “Cognitive Aspects of Number Notation and Arithmetic Algorithms” (supervisors: Prof. Kai-Uwe Kühnberger and Prof. Inge Schwank). Institute of Cognitive Science, University of Osnabrück.

ACADEMIC AND PROFESSIONAL CAREER

2015	Organiser of the Symposium <i>Oswald Wiener. Selbstbeobachtung – Denkpsychologie</i> (Symposium on Introspection and Psychology of Thought). kunsthau muerz. Müzzzuschlag (with Dr. Thomas Eder and Cornell Schreiber MSc.)
2012	Teaching Assistant at the University of Osnabrück (with Dr. J. Griego), Cognitive Psychology and Neuropsychology
2011	Co-Organiser of the Symposium <i>Oswald Wiener. Innenschau – Zusammenschau</i> (Symposium on Introspection and Psychology of Thought). kunsthau muerz. Müzzzuschlag (with Dr. Thomas Eder)
2010–11	Associated Member of Study Project “COUNT” in the Cognitive Science Master programme at the University of Osnabrück (supervised by Prof. Kai-Uwe Kühnberger)

CONFERENCE CONTRIBUTIONS (selection)

2016	<i>Tagung experimentell arbeitender Psychologen 2016 (Conference of Exp Psychologists)</i> University of Heidelberg, Poster (w. C. Schreiber & S. Schneider): “Representational Dynamics and Problem Spaces: An Exploratory Case Study”.
2015	<i>Oswald Wiener. Selbstbeobachtung – Denkpsychologie. Symposium on Introspection and Psychology of Thought.</i> kunsthau muerz. Invited speaker: “Entwicklung eines Versuchsdesigns für die experimentelle Selbstbeobachtung”. Austria. Oct 8–11.
2014	<i>Interdisciplinary College (IK) 2014</i> , Spring school in Günne, Lake Möhne Germany. Poster contribution (with Stefan Schneider): “Becoming an Expert in Mental Paper Folding”.

MEMBERSHIPS

Gesellschaft für Kognitionswissenschaft e.V. (GK) (student member)
Artificial General Intelligence Society (full member)

PUBLICATIONS

Angerer, B. (2015). Orientierungsaufbau in einem ungewohnten Zahlenraum. In: Raab, T. & Eder, T. (Eds.). *Selbstbeobachtung: Oswald Wieners Denkpsychologie*. (pp. 237–258). Berlin: Suhrkamp.

Angerer, B. & Schneider, S. (2013). General Problems of Unified Theories of Cognition, and Another Conceptual Commitment of LIDA. [Peer Commentary on “Conceptual Commitments of the LIDA Model of Cognition” by Franklin, S. et al., same issue, pp. 1–22.] *Journal of Artificial General Intelligence*, 4(2), pp. 26–28.

Schneider, S., Abdel-Fattah, A. M., Angerer, B., & Weber, F. (2013). Model Construction in General Intelligence. In: *Artificial General Intelligence – 6th International Conference, AGI 2013, Beijing, China, July 31 – August 3, 2013 Proceedings*, LNCS 7999, pp. 109–118. Springer: Berlin Heidelberg.

DECLARATION

I declare that the thesis *Representational Dynamics in Problem Solving* was composed by myself and that the work contained is my own except where explicitly stated otherwise.

Development of the task, study design and work on foundational and methodological questions (discussed in Chapter 2, but including work relevant for the whole thesis) was conducted in close and equal cooperation with Cornell Schreiber MSc at the University of Vienna. A paper based on the results of this thesis, with Mr. Schreiber as its co-author, is currently under review (Angerer & Schreiber).

This work has not been submitted for any other degree or professional qualification.

Vienna, March 2017

Benjamin Angerer

ACKNOWLEDGMENTS

I wish to thank Cornell Schreiber in close cooperation with whom the study presented in this thesis was developed, and Stefan Schneider who provided encouragement and significant help with data analysis. Additionally, both of them provided valuable comments on my ongoing writing without which the quality of this thesis would have suffered greatly.

I am grateful to Moritz Meier and Moritz Jacobs who helped with the software side of things. I am also grateful to Clemens, Ferdi, Maike and Fam who contributed considerably through countless conversations about methodology, epistemology and cognitive psychology.

I wish to thank my supervisor Prof. Markus Peschl who always encouraged me to follow my ideas. Furthermore, I owe gratitude to Dr. Paolo Petta and Prof. Urban Kordeš who provided support and advice beyond their call of duty.

I am also very thankful to all study participants who voluntarily offered considerable amounts of their time and mental effort to provide us with valuable empirical data.

Finally, I wish to thank Oswald Wiener, Tom, Thomas, Mike and Johannes whose ideas were a significant influence for all of my work in the last years. I am looking forward to future discussions and cooperation.

CONTENTS

1	INTRODUCTION	1
1.1	Human problem solving	1
1.2	Problem solving à la Newell & Simon	3
1.3	Different problem spaces for different purposes	7
1.3.1	One space to rule them all?	7
1.3.2	Different spaces for the same problem	9
1.3.3	Changing spaces over time	9
1.3.4	Sudden changes	11
1.3.5	Multiple problem spaces	13
1.3.6	Problem spaces in Soar	15
1.4	Where do problem spaces come from?	16
1.4.1	Reasons for the de-emphasis of problem understanding	17
1.4.2	The methodological dilemma	19
1.5	Summary	19
1.6	Overview of the remaining thesis	20
2	DESIGN AND METHOD	21
2.1	Design: Task properties	21
2.1.1	Existing approaches: Insight problem solving	21
2.1.2	Existing approaches: Complex problem solving	22
2.1.3	Task properties	24
2.2	Design: Key properties of the study	24
2.2.1	Exploratory single-case studies	25
2.2.2	Microgenetic studies	26
2.2.3	Acquiring process-level data	27
2.2.4	Mental imagery	30
2.3	Design: Overview of the key properties	32
2.4	Method: Overview of the study	32
2.5	Method: Tasks	33
2.5.1	Training tasks	33
2.5.2	Main tasks	35
2.5.3	Overview of sub-tasks	36
2.6	Method: Subjects	37
2.6.1	Finding and selecting subjects	37
2.6.2	Pilot cases	38
2.7	Method: Instructions	38
2.8	Method: Procedure	41
2.8.1	Training session	42
2.8.2	Main sessions	43
2.9	Method: Setup	44
2.9.1	Room	44
2.9.2	Equipment	44
2.9.3	Software	44
2.9.4	Gathered data	45
2.10	Data analysis: Theory	45
2.10.1	Trace analysis	46
2.10.2	Competitive argumentation	47
2.10.3	Microgenetic learning analysis	48

2.11	Data analysis: Experiences from the pilots	49
2.11.1	Segmentation	49
2.11.2	Incorporation of further data sources	50
2.11.3	Annotation	50
2.11.4	Cooperative interpretation	51
2.11.5	Dealing with the subjects' private epistemologies	51
2.11.6	Emergent annotative codes	52
2.11.7	Example 1: Codes for coarse cognitive activity	53
2.11.8	Example 2: Report-type codes	54
2.11.9	Summary	56
2.12	Conclusion	56
3	TASK ANALYSIS	57
3.1	Related research	59
3.1.1	Mental imagery	59
3.1.2	Paper folding	60
3.1.3	Cognitive skill acquisition	61
3.2	An overview of the cross-folding domain	62
3.2.1	The basics: Notation and varieties of folding	62
3.2.2	The primitive forms: The V and its relatives	64
3.2.3	Sheet size, fold thickness and aspect ratio	66
3.2.4	Chirality	67
3.2.5	Out-folding and in-folding	68
3.2.6	Folding procedures and their relation to fold states	71
3.2.7	An overview of variants and common abstractions	76
3.2.8	A general, recursive procedure	77
3.2.9	Barbra notation	78
3.2.10	A recursive algorithm	79
3.2.11	A simpler, iterative procedure	80
3.2.12	A mathematical description of NV and DV	84
3.2.13	Conclusion	84
3.3	Observations from the pilot studies	86
3.3.1	Repetition, confusion and mind-wandering	86
3.3.2	On the use of folk-psychological terminology	87
3.3.3	Overview of sub-tasks	87
3.3.4	Initial setup – Associations, actions and motor imagery	88
3.3.5	First drawings	92
3.3.6	Dealing with complexity: First abstractions	95
3.3.7	Systematising sketch generation	101
3.3.8	Beginnings of a recursive procedure	106
3.3.9	Conclusion	108
3.3.10	Overview of some possible sub-tasks	108
3.4	Possible representational-level problem spaces	110
3.4.1	Sheet space	111
3.4.2	Sketch space	115
3.4.3	Rule spaces	119
3.4.4	Meta-cognitive spaces	120
3.4.5	General meta-cognitive space	121
3.4.6	Report space	121
3.4.7	The whole layout	122
3.4.8	Layout variants	122
3.4.9	Concerning representational-level problem spaces	124
3.5	Conclusion	124

4	DISCUSSION	127
4.1	Pointers for theory formation	127
4.2	Future work	130
4.2.1	Lessons from the pilot studies	130
4.2.2	Further questions	132
4.3	Conclusion	133
	BIBLIOGRAPHY	135
A	APPENDIX A: STUDY MATERIAL	149
A.1	Consent form	149
A.2	General instructions	149
A.3	Protocol instructions	150
A.4	Shortened instructions (Reminder)	151
A.5	Tasks	151
A.5.1	Training session	151
A.5.2	Session 1	152
A.5.3	Session 2	154
A.5.4	Session 3	155
A.5.5	Session 4	156
B	APPENDIX B: STUDY GUIDELINES	157
B.1	Training session	157
B.1.1	Welcome & explanation of the session	157
B.1.2	Consent form	157
B.1.3	General instructions	157
B.1.4	First task block	157
B.1.5	Debriefing for first block	159
B.1.6	Second task block	159
B.1.7	Debriefing about second block	161
B.1.8	Final conversation	161
B.2	Main sessions	162
B.2.1	Welcome	162
B.2.2	Instructions	162
B.2.3	Selection of tasks	163
B.2.4	Debriefing interview	163
B.2.5	Closing the session	163
B.2.6	Final remarks	163
C	APPENDIX C: SOFTWARE	165
C.1	Study software	165
C.2	Folding implementations	166
D	APPENDIX D: ABSTRACT	170
D.1	Abstract	170
D.2	Zusammenfassung	171

INTRODUCTION

“(too many toy tasks?)”

— Newell 1979, p.1

1.1 HUMAN PROBLEM SOLVING

Problem solving is a complex human activity usually seen as the hallmark of “higher cognition”, making use of the various faculties of our cognitive architecture. The breadth of the definition of what constitutes problem solving varies greatly, but the general consensus is that it is concerned with *multiple-step goal-oriented activity* with a usual duration of *several minutes to several hours* (Burns & Vollmeyer, 2000; VanLehn, 1988). It comprises a *task*, i.e. an initial situation together with a desired situation or *goal*, an *environment* in which the task is to be acted upon, and the construction of a solution proceeding from the knowledge, skills and tools available to the problem solver.

This very general and abstract definition already hints at the fact that problem solving is not a very specific kind of activity. Rather, we can understand it as all those activities we engage in when working towards a desired situation. The broad range of activity this might involve is reflected in research on the topic, which includes inter alia: Deduction (Schwartz & Fattaleh, 1972), induction and abduction (Cifarelli, 1999; Schmid, 2005; Simon & Lea, 1974), analogy (Chan et al., 2012; Gralla et al., 2012), mental imagery (Clement, 2008; Qin & Simon, 1992a), and meta-cognition (Anderson & Fincham, 2014; Berardi-Coletta et al., 1995). Even though the majority of research on problem solving has traditionally assumed a rather decoupled, purely cognitive perspective, in recent years research has also started to include aspects from embodied and situated approaches, such as studies emphasising interaction with the environment (Kirsh, 2009a), and social interaction (Clancey, 2006; Dunbar, 1997). In a fashion similar to these more modern approaches, this thesis attempts to shed light on aspects of human problem solving classical cognitive science tended to overlook.

Historically, the study of human problem solving in cognitive science can be traced back to several important influences, most notably *Gestalt psychology* (Duncker, 1935; Wertheimer, 1959) and the Würzburg school’s *psychology of thought* (Bühler, 1907; Külpe, 1912; Mandler, 2011; Selz, 1913). Through de Groot’s “Thought and Choice in Chess” – a verbal protocol study of Chess players building on Otto Selz’s theory of productive thought – these early cognitive theories reached North America, and in particular Allen Newell and Herbert Simon (Baars, 1986). In the early 1950s Newell & Simon started investigating human problem solving with protocol studies themselves, with the crucial addition of concepts from information and computer science (de Groot, 1965, p.376f; Newell et al., 1958). This work culminated

in Newell & Simon (1972), describing a theory of problem solving which is used to this day and has remained remarkably unchanged¹.

This thesis aims to be a thorough and critical exploration of Newell & Simon's theory with special regard to aspects of human problem solving it tends to de-emphasise.

To this end, we will:

- (1) Analyse the current state of the *theory* of problem solving and identify its strengths and weaknesses,
- (2) Present a new *study design* and *task* (iterated, mental paper folding) designed to advance theory formation, and finally
- (3) Present a *task analysis*, based on insights from pilot studies and formal investigations of the task's domain

The main question we will derive from our analysis of problem solving theory – and which the study we will present later is supposed to shed light on – is:

How are the cognitive representations used by a problem solver constructed, and how do they change over time?

These aspects of problem solving, such as set-up and choice of representation, potential changes within one representation, switching between different forms of representation were roughly summarised in the thesis title as *representational dynamics* (cf. Dietrich & Markman, 2014).

Since *representation* is a central term in Newell & Simon's theory, their theory making it very easy to use the term in an objectifying manner, and since the term is not without controversy in cognitive science (cf. e.g. Bickhard & Terveen, 1996; Brooks, 1991), we have to say a few words regarding our use of it, before we can proceed to introduce Newell & Simon's theory.

Regarding the use of the term 'representation' in this thesis

The notion that a cognitive agent possesses internal states which are manipulated by cognitive processes and which thus mediate between the agent's sensations and actions is usually considered of the central tenets of cognitive science (Friedenberg & Silverman, 2011, p.3f; Thagard, 2005). These internal states are commonly called (mental or cognitive) *representations*.

Newell (1990, p.59f) gives a typical account of this perspective, by explaining that something is a representation of something else as long as there are processes which maintain some sort of *lawful relation* (not necessarily correspondence) between the represented thing and the representation. This representation can then be manipulated by the representing system in order to reason about the represented thing, even in its absence. Since this is a very generic statement, there might be many different ways in which such a lawful relation could be fulfilled (Newell, 1990, p.61).

In the context of a cognitive agent acting in an environment, this lawful relation might be maintained by the interplay of perception and actions affecting the agent's environment, although we should note that Newell

¹ Although it has not remained without criticism (e.g. Goel, 2010; Kirsh, 2009a; Wiener in Simon, 1994).

explicitly mentions that neither does the representing system have to be necessarily internal to an agent, nor does the represented situation have to be external. Hence, we could also consider e.g. “nested thoughts” which referring to abstract states of affairs (everything being internal to the agent) as representations in the above sense, or the notebook a forgetful person might use to aid their memory (external representation).

Even though this generic formulation does by itself not entail everything which some of its philosophical opponents have brought forth against it, there have been many valid and important criticisms of specific, narrower versions of the general idea (e.g. of actual proposed models), these include:

Questioning the implied sequentiality of a perceive-think-act cycle (Hurley, 2001), the role of direct interaction without detailed representations of the environment (Agre & Chapman, 1990; Brooks, 1991; Clark, 2008), the observation that while phenomenally it may seem to us as if we perceive a rich environment (the “grand illusion”) available information is often meagre and better described as indexical/deictic (Agre, 1988, chap.C; Ballard et al., 1997; Noë & O’Regan, 2000), and challenging theories of representation with ontologically dubious requirements towards the relation between representation and represented state of affairs, such as truth-preserving correspondence (with representations being “copies” of objective states of affairs in the environment) (Glaserfeld, 1987).

Additionally, many arguments have been brought forward against specific formalisms for describing cognitive representation, such as amodal symbol systems (Bach, 2008; Barsalou, 1999), connectionism (Fodor & Pylyshyn, 1988; Minsky, 1991), and probabilistic approaches (Wiener, 2015).

The use of representation adopted in this thesis does not assume any of these specific perspectives, but rather refers to the broad definition given at the outset of this section: *We call something a representation if we might assume there is some process allowing us to refer to a represented state of affairs even in its factual absence.* This is regardless of whether this representation is supposed to be amodal or perceptual, abstract or embodied, symbolic or distributed, and truth-preserving or subjectively constructed. With regard to the cited criticisms of representational theories, we go along with Markman & Dietrich (2000) who propose a reconciliatory perspective: The modern approaches such as situated and embodied cognition and constructivist theories raise important challenges and make valuable contributions to how a useful and workable theory of representation must look like, but it would be unhelpful to discard the concept altogether.

Having made these preparatory remarks, we continue with an introduction to the classical problem solving theory by Newell & Simon².

1.2 PROBLEM SOLVING À LA NEWELL & SIMON

Roughly, Newell & Simon’s theory proposes that a problem solver’s understanding of

- (1) their current situation,
- (2) their desired situation, and

² The main reference for the whole of Section 1.2 is Newell & Simon (1972), hence we will only cite it if quoting specific passages.

(3) possible actions changing this situation

can be described in terms of a mental space, which they called a *problem space*. This internal problem space describes how a subject understands possible actions in and parts of their environment relevant to the problem such as an external observer would see them – the *task environment*³. Newell and Simon imagined a problem space as a graph-like structure, where the situations a problem solver is considering are the nodes, connected by arcs which stand for the solver’s actions transforming one situation into another, in their terms (Simon & Lea, 1974, p.109f):

A problem space consists of elements called *knowledge states*. Generative processes called *operators* can take such states as input, and in turn produce a new state as output.

The initial situation, the goal situation and any situation in between which we might be in or are considering should be expressible as knowledge states, while the operators correspond to any state-modifying action we might engage in. If an action transforms one state into another, those two states are adjacent. Trying to solve a problem can be described as *searching* for a continuous chain of operators which constitute a fitting connection between the current state and the goal (Newell & Simon, 1976).

Problem spaces: A toy example

In this section, we will consider a simple toy problem and ask the question how a problem space would have to be constructed such that we can describe our solving of the problem as working within such a space.

Consider this simple blocks world problem (after Laird, 2012, p. 47f)⁴:

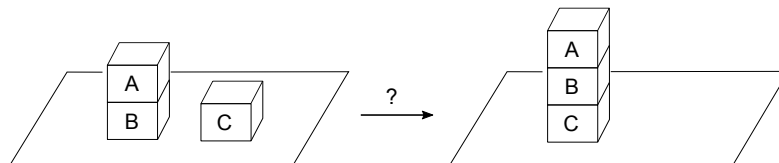


Figure 1: A simple blocks world problem. After Figure 3.2 from Laird (2012, p.47).

How can we transform the situation given on the left into the right one? First, we have to come up with a way of representing the *states* of the blocks world. Leaving aside the exact physical layout of the situation, we can for instance describe a situation by stating which block is stacked on which other block, or on the table. Further information, like which blocks there are overall, or the question if a block’s top is free or whether it has another block stacked on it can be *derived* from this formulation and does not have

³ Note that Newell & Simon define task environment as something objectively given through the problem – without saying much about how this objective judgement is made. For many problems, it might be hard to decide which parts of the environment could turn out relevant for its solution and which could not. Still, the concept remains useful, in capturing the observation that for many problems we can abstract from this or that aspect of our environment: A game of chess remains the same game even if would have been played in another room.

⁴ The “blocks world” is a widely used task domain in AI research. Most notably it was used in SHRDLU, an early natural language processing and reasoning system (Winograd, 1972).

to be stated explicitly⁵. In the initial state, blocks b and c are lying on the table and block a is lying on top of block b. So, we could write down the first state like this⁶:

$$S_0 = [\text{on}(a,b), \text{on}(b,\text{table}), \text{on}(c,\text{table})]$$

In turn, the goal state on the right would look like this:

$$S_G = [\text{on}(a,b), \text{on}(b,c), \text{on}(c,\text{table})].$$

What about the actions with which we could manipulate the states? The only action usually considered in blocks world problems is *moving* the blocks around, so our only operator is:

$$\text{move}(\langle \text{block} \rangle, \langle \text{to} \rangle),$$

where $\langle \text{block} \rangle$ is the block to be moved and $\langle \text{to} \rangle$ is its destination (the table or another block). We are not done yet, however: The move-operator has important *constraints* which must be fulfilled for it in order to be applicable – the block to be moved and its destination cannot be covered by another block. Furthermore, we cannot move a block onto itself⁷.

Figure 2 shows the expansion of the problem space for this problem. Each rectangular node represents a blocks world state and each arrow is the corresponding move-operator:

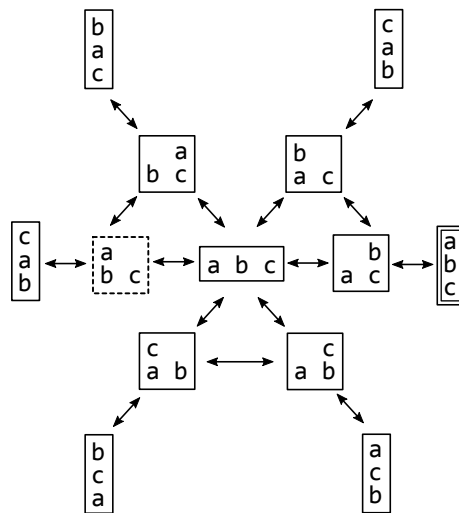


Figure 2: Visualisation of the blocks world problem space for the problem in Figure 1. The initial state is dashed, the goal state is double-lined.

With the concepts introduced, we can understand trying to solve such a problem, as *searching* the problem space. Since we can move blocks around as we please – even move a block back to where we just took it from and

⁵ Additionally, if the problems we wanted to pose/solve required it, we could introduce further notions, such as which blocks are left/right of which others, or which blocks statically support each other.

⁶ The way in which we have written down the states here looks similar to list notations and first-order predicate logic such as is often used for problem solving systems in AI. Note however that we are free to use any other form of representation e.g. an iconic depiction of the situation, as long the processes operating with these representations can deal with the form we have chosen.

⁷ It may sound very artificial to require this explicitly. However, when formalising a problem introducing such artificialities often becomes necessary to ensure a model will behave correctly.

repeat this ad nauseam – there is an *infinite number of potential solutions* for our example problem. Even so, this problem is hardly difficult for people and they can find a good solution easily. This is due to two further properties of human problem solving we have not touched upon yet:

First, with many problems, people not only try to solve them in any way, but they have *additional criteria* with regard to the kind of solution they want to find, which they use to *guide their search* of the space. For instance, in many cases people try to find the solution with the shortest amount of steps (alternatively: the cheapest, in game playing often the one which maximises a score, the most ecologically friendly, etc.). Concerning our toy problem this means that if we are interested in finding a short solution, we would very likely never consider making a move and reverting it immediately afterwards. Secondly, people do not search a problem space blindly, but strategically: They use their knowledge about the problem to *evaluate* the relative merits of possible actions and decide on the one they deem more promising. Since such knowledge rarely directly yields a solution, but rather *informs* the search for one, this process is called informed or more commonly, *heuristic search* (Newell & Simon, 1976).

The shortest solution to our example problem would be the following ordered sequence of 3 operators:

Solution = [move(a,table), move(b,c), move(a,b)].

Since operators move through the problem space *from state to adjacent state*, we speak of such sequences as *paths* – a path leading from an initial state to a goal state is hence often called a solution path. Note that states are only adjacent if there is an action which transforms one into the other. If we want to describe someone's behaviour in terms of a problem space and at some point they seem to move from one state to a non-adjacent one, then something has happened which was relevant to the problem, but could not be expressed in the problem space we assumed, i.e. we defined this problem space incorrectly and have to introduce an action which connects those two states.

In large parts of this thesis, the notion of problem spaces will be used as a tool for us researchers to *describe and analyse another person's thinking*. In this situation, we must not forget when thinking about paths through a problem space that this "global perspective" on the space is something that we are imposing on the subject *after the fact* (cf. Newell, 1967, p.49). Owing to their limited *working memory capacity* (Baddeley, 2007; VanLehn & Ball, 1991, p.149), in any given moment a problem solver will very likely only have the current state⁸ and maybe a very small number of potential successor states in mind (Newell, 1990, p.100), marking the difference between *immediately available* and *derivable knowledge*.

In Agre (1994, p.423), Newell proposes the following analogy: When navigating through a city, a person necessarily has a very limited perspective, namely whatever they are able to see from their current position. When we want to understand their movement through the city, we can describe their movements relative to a map of the city, but using the map is our privilege – the person finding their way through the city is always bound to their current, partial perspective.

⁸ Note that there is always just one current state, and while many operators *can* potentially be applied, it is always only one which will be applied (Newell, 1979, p.7).

It is important to remember the distinction between immediately available and derivable knowledge. If it was true that being able to generate a state in principle and possessing full knowledge of a state are the same, knowing the rules of chess would be same as having mastered the game.

1.3 DIFFERENT PROBLEM SPACES FOR DIFFERENT PURPOSES

Even though the central concepts of the theory of problem solving have not changed substantially since 1972 (cf. Ohlsson, 2012b; VanLehn, 1988), it is important to note that the concept of the problem space has evolved several somewhat different usages within different research contexts in cognitive science. Since the problem space is one of the main notions of the theoretic parts of this thesis, the following section will try to carefully distinguish between these uses. In doing so, we will also examine the *explanatory scope* of the concept of problem spaces with respect to the psychology of problem solving.

In particular, we will distinguish several levels of analysis on which problem spaces can be employed to describe problem solving activity: In analogy to David Marr's levels of analysis for cognitive processes, we distinguish between (1) spaces which are used as generic, descriptive devices for problem-solving behaviour (computational level), (2) spaces in which we can describe the specific way in which someone thinks about a problem (representational level), and finally (3) spaces which are actually meant to instantiate the problem-solving process (implementational level) (Marr & Poggio, 1976).

1.3.1 Computational level: One space to rule them all?

The first important question to be asked is who or what does a problem space pertain to? If a problem space is an analytical property of a *task domain*, then we can describe everyone attempting to work at the same problem as traversing the identical problem space. But if the problem space is unique to the domain and not to a *problem solver*, we can only use problem-solver independent domain properties to define the space. Often, these domain properties correspond largely with the *task environment*, i.e. the area in which the problem solver's activity is supposed to lead to a solution.

While in principle it is possible to exemplify this difference in the blocks world we introduced above, it is easier to see in more complex task domains. In chess, for instance, the states of such a space would be the position of all pieces on the board and the operators would move a piece constrained according to the rules of the game. In this space, we could then describe all moves actually made in a game. Beyond that, we could also describe all moves *considered* by a player, if we can somehow acquire those data. Theoretically, it would be easy enough to make an artificial agent print out all the moves it is considering, but for a human player, we would have to rely on an external measure – such as having them *verbalise* which moves they are thinking about⁹ – introducing differences into our analysis which are not necessarily visible on the chess board.

⁹ In fact, *verbal reports* of someone's thinking, gathered in a so-called *protocol*, will be the main source of empirical data this thesis is concerned with.

It gets more complicated however, when we do not just want to describe a game, but also *explain* why a certain path through the space has been taken. In order to do so, we need more information than which moves were considered and subsequently taken. If we want to model how the player *decided* which move to make, we need to include whatever information they used to make this decision. This could for instance include intermediate goals they have formulated, criteria they used to evaluate possible moves, and how they modelled the possible intentions of their opponent.

Moreover, we know from psychological research on chess experts that their thinking differs more fundamentally from how novices approach the game: Not only might they use different intermediate goals and evaluation criteria, but they might employ *other kinds of representations* for board states altogether. Such advanced representations might group together many board states based on perceptual and maybe also abstract similarities (“chunks”) (Chase & Simon, 1973; León-Villagrà & Jäkel, 2013)¹⁰. If we were to describe such an expert’s considerations in the “objective” problem space, their movement through the space might look confusing to us: Since they do not necessarily consider single board states (as the objective space does), but groups of states, clustered based on individual similarity criteria, the expert’s progress might look discontinuous to us – as if they were “jumping” between distant areas of the problem space. In order to explain such jumps, we have to refer to the representations the expert was actually using – which means, we have to include whatever they are into the problem space and change our definition of what constitutes a state accordingly.

So, by including the specifics of a subject’s or a group of subjects’ thinking, we can increase the “psychological accuracy” of our problem space, but we are trading it off against universality. And vice versa: While the objective problem space has the advantage of being universally applicable, we can only describe *what* problem solvers are doing or considering, and are very limited with respect to the question of *how* this behaviour came about. These objective kinds of spaces are sometimes called the *minimal*, *legal*, or *basic problem space* of a problem, emphasising the view that while more advanced spaces people use might *add* structure to the minimal space, they would still include it (Newell & Simon, 1972, chap.5; Newell, 1979, p.5). However, since we want to emphasise the coarse, descriptive level of analysis assumed when using such spaces, we will call them *computational-level problem spaces*, referring to David Marr’s distinction between levels of analysis of cognitive activity (Marr & Poggio, 1976).

It does not really make sense to attribute computational-level problem spaces to a person, but they can be very useful in other ways: Since they correspond largely to a straight-forward representation of the external task environment plus those operators defined as legal in a formal problem definition, we can discuss them *in contrast* to a subject’s actual problem representation: How do they think about it and how does their thinking relate to our strict definition of the problem?

Therefore, we often find computational-level problem spaces in abstract discussions of tasks (task analyses), where a problem can be discussed without having to address the possible idiosyncrasies of actual people working on the task (e.g. Newell & Simon, 1972, p.144f). Additionally, these kinds of

¹⁰ Although its low complexity does give neither much opportunity or need for abstraction, even in our blocks world example above we could e.g. start thinking in terms of more complex, composite objects such as towers or arches.

spaces are used where we are not interested in the specifics of how humans approach a problem, i.e. in result-oriented AI research¹¹ (e.g. Pang et al., 2004).

It is in these kinds of contexts (abstract discussions of tasks and AI) that we often encounter phrases like “the problem space of problem X”, implying that a problem space is fully determined by the problem itself. In such contexts – i.e. those where we do not claim that this space is the cognitive representation a person uses, but only an abstract description of a how to conceptualise manipulating the task environment – this use makes perfect sense. But since it is an easy way of talking about problems it sometimes also occurs elsewhere in problematic contexts, such as instructional material in cognitive psychology (e.g. Goldstein, 2011, p.333f)¹². It is the appearance of computational-level spaces in contexts where the implication might be that they also are what people have in their heads, that may unwittingly leave the misleading impression that problem spaces *in general* are objective properties of a task domain and not the *subjective mental constructs of individual problem solvers* (Simon, 1991, p.36).

1.3.2 Representational level: Different spaces for the same problem

But there is not just a choice between completely universal computational-level spaces or *representational-level problem spaces* fitted to describe an individual. For many tasks with a clearly structured task environment, people can be described quite accurately as working in one out of a small set of predictable problem spaces. In fact, a lot of research has been conducted regarding the question of what the initial choice of problem space depends on, e.g.: Different instruction wording and problem isomorphs (presenting a logically identical problem with different “cover stories”) (Hayes & Simon, 1974; Simon & Hayes, 1976), the influence of preceding mental activity (Schunn & Dunbar, 1996), scaffolding of the task environment (Kirsh, 2009b), or domain-specific expert knowledge (Nokes et al., 2010).

For instance, Chi et al. (1981) compared how novices and experts categorised physics problems, showing that while experts construed problems to be about *domain principles*, such as Newton’s second law, novices categorized the problems according to *surface features* such as problems involving inclined planes or pulleys. These different categorisations in turn suggest different ways of representing the problem, as well as different actions one might try – in short: Different problem spaces.

1.3.3 Representational level: Changing spaces over time

In addition to differences between groups of people there also can be differences in the same group of people (or within the same individual) *at different points in time*, or in other words: The problem spaces in which people work *might change* with experience or people might even *switch* to working in a qualitatively different space. Depending on the qualitative nature of these

¹¹ We specifically qualify the kind of AI we mean here as “result-oriented”, since there is actually a lot of AI research which takes human cognition as inspiration or even aims to explicitly model it.

¹² And also in Figure 2!

changes as well as their time scale, they are attributed to be the effects of *practice*, *learning*, or *development*. This is why the range of “several minutes to several hours” cited in the beginning is such an important part of the consensual definition of problem solving: If we looked at larger time frames we cannot assume problem spaces to be static any longer, since due to learning from experience they would change over time.

Even when looking at problems which are pretty straight-forward to understand, like *cryptarithmic problems*, we can find that people are using different problem spaces or change the one they are employing over time. A cryptarithmic problem is a recreational maths puzzle, where one has to assign digits to letters in order to make an arithmetic expression true, for instance:

$$\begin{array}{r} S\ E\ N\ D \\ +\ M\ O\ R\ E \\ \hline M\ O\ N\ E\ Y \end{array}$$

To solve this problem, we have to assign a unique decimal digit to each of the letters D, E, M, N, O, R, S, and Y. The presentation of the problem akin to a pen and paper arithmetic task suggests dealing with the problem in terms of columns and carries. If we look at the example problem, we can see quite easily, for instance, that a carry must have taken place in the left-most column, since both addends are 4 digits long, but the result has 5 digits.

In their task analysis of cryptarithmic, Newell & Simon describe several different problem spaces (Newell & Simon, 1972, chap.5): A basic problem space (see discussion above), an augmented problem space, which adds concepts like carries, disjunction, digit parity and inequality, and an algebraic space which allows arithmetic manipulation of “letter equations”. In (ibid., chap.6) they proceed to infer from the protocols of an individual’s verbal reports that he was using a variant of their augmented problem space for the duration of the study (cf. also Newell, 1967, p.32f).

But what would happen if Newell & Simon had enlarged the time frame of their study and let the same subject solve several cryptarithmic problems for, say, a period of several days? If they had continued to collect verbal reports of his thinking, they might have found evidence for *learning* in their protocols – and the subject might have changed to a more advanced, algebraic problem space.

So, if we want to keep using the concept of problem spaces for such situations, we have to concede that they can *change gradually*, as experience with a task domain teaches us to better classify states, or suggests the introduction of new, useful operators. Changing problem spaces in turn might allow us to traverse them differently and hence cause a change of solution strategy:

In the “Tower of Hanoi” puzzle – one of the most widely used tasks in classical problem solving research – we have to move differently sized disks between three pegs. The difficulty is that we are only allowed to stack smaller disks onto larger ones, but not the other way around (Lucas, 1979). Typically, subjects start out with a trial-and-error strategy, but once they developed the ability to see partial disk stacks as “sub-towers” they can in turn develop a more effective, recursive solution strategy (Anzai & Simon, 1979; Ruiz & Newell, 1989).

Allowing problem spaces to change might at first seem to weaken the concept, since we have to stop thinking of them as the static, unchangeable “grammar” of a person’s perspective onto a problem. On the other hand, this “weakening” allows us to use (changing) problem spaces to describe problem solving over longer stretches of time, hence enlarging the scope of the concept.

Yet, up to now, we can only *describe* the problem solving as taking place in a changing space, we cannot *explain* from within the space how these changes came about. Instead, we have to assume that *additional processes* must be at work which mediate this change, such as *learning*.

1.3.4 Representational level: Sudden changes

Changing one’s problem space is not solely a function of accumulating experience. Sometimes changes can occur quite suddenly and thoroughly. There is a class of problems called *insight problems*, which are renowned for subjects “getting stuck” in them, followed by a sudden moment of insight – an “Aha!” moment (Bühler, 1908; Chu & MacGregor, 2011). A widely used explanation for these long periods of unsuccessful rumination (*impasses*) is that people start out representing the problem in the “wrong” problem space – one which does not include the solution. The trick to solve the problem is not to excessively search this initial problem space, but to *re-represent* the problem into another problem space, one that does include the solution (Knoblich et al., 1999; Ohlsson, 2011)¹³:

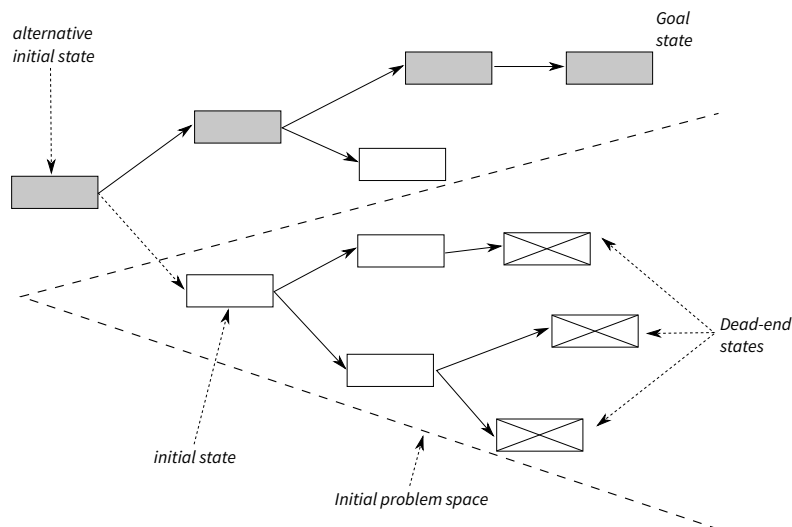


Figure 3: Illustration of an initial problem space not containing the goal state, and an alternative, less constrained space which does include it. After Figure 4.3 from Ohlsson (2011, p.105).

A classical example for these kinds of problems is Duncker (1945)’s candle problem (Figure 4): The task is to fix a candle to a wall and light it, given these materials: The candle itself, a book of matches and a box of tacks. The

¹³ As with the gradual changes through learning discussed above, sudden changes brought about by insight can be described in terms of changing problem spaces, but in order to explain them, we have to assume additional processes outside of the problem which bring about this change.

canonical solution to this problem is to fix the box the tacks were in to the wall using a tack and putting the candle in it. Many people have trouble finding this solution, since they do not consider the box as a separate object to be used, but only as the container for the tacks (Weisberg & Suls, 1973). Duncker termed this phenomenon of not being able to change the default role one considers for an object *functional fixedness*.

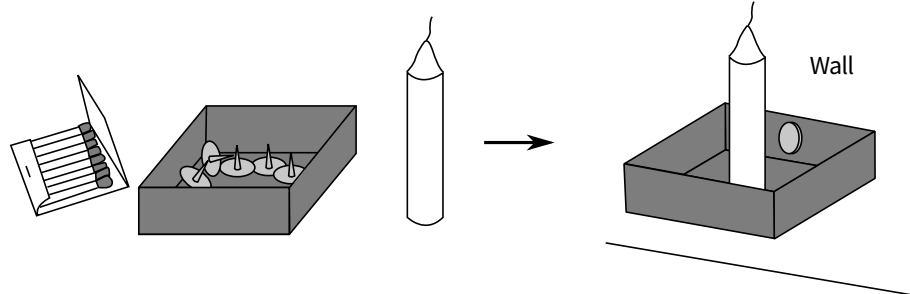


Figure 4: Duncker (1945)'s candle problem: By implicitly categorising the box of tacks as not as an active ingredient for a solution, many subjects fail to see the solution shown on the right.

In problem space terminology, the impasse in the initial problem space for the candle problem comes about because the tack box is represented as container for the tacks. Therefore, no amount of search in this problem space will yield the solution of using the box as container *for the candle*. A possible re-representation would be to represent the tack box as a *generic* container, allowing it to contain any other objects which fits in it. This is a good description of the problem spaces a person might be using to think about the problem, but if we want to understand the nature of insight problem solving, the remaining question is *how this change of problem spaces is being brought about*.

Before we approach this question, let us summarise the distinctions introduced thus far: Problem spaces can be used as very general descriptive frameworks of people's activity in a task environment (computational-level problem spaces), or less general in order to explain the specific way a group of people or an individual is thinking about a problem (representational-level problem spaces). Therefore, the appropriate choice of problem space for analysing someone's problem-solving behaviour does not only depend on the task and the individual's representations, but also *on the aim of our analysis*.

The specificity of representational-level spaces means that there is a multitude of possible problem spaces for every problem. Additionally, spaces are subject to change, gradually as a function of experience (learning), as well as suddenly due to insight. There is a trade-off between the universality and "psychological accuracy" of problem spaces.

1.3.5 Explaining change: Multiple problem spaces

While we can *describe* all the phenomena mentioned above as specific problem spaces, Newell and Simon's original problem space theory does not make any predictions of what happens beyond the problem space one is currently searching. Hence, like we have alluded to in the example with the candle problem, phenomena like learning, re-representation, or more generally, changing or switching problem spaces are outside of its *explanatory scope*.

We already noted that if we wanted to explain these phenomena, we would have to assume additional processes which instigate the changes in our current problem space. One interesting proposition in problem solving research is that instead of coming up with a completely different theory, we can describe these additional processes as taking place in a problem space themselves – a second problem space which works “above” the first one, but instead of searching through possible actions affecting the state of our problem, it searches through possible changes of the first problem space¹⁴.

The first time such a multiple problem space-architecture was suggested, was in Simon & Lea (1974)¹⁵. Suggesting rule induction tasks as an example, they proposed that subjects solve these by searching two distinct spaces: A *rule space* in which possible rules underlying the given problem are being searched, and an *instance space* in which these rules are employed and put to the test.

Similar ideas were explored in Klahr & Dunbar (1988) where the authors investigated scientific reasoning (understood as a form of inductive problem solving). Their “Scientific Discovery as Dual Search” model (SDDS) proposes that scientific reasoning takes place in a *hypothesis space*, where hypotheses are formulated based on current evidence and an *experiment space*, where experiments are designed and carried out, yielding new evidence.

A further expansion of these models have been proposed in Burns & Vollmeyer (2000) who add a third space on top of Simon and Lea's dual space hierarchy. In this so-called *model space* different possible task understandings are searched. Reasoning in model space is supposed to reflect ambiguous mappings of the task instruction to a subject's understanding of the task and possible re-representations of this understanding when it passes in the lower spaces “escalate” up to model space. This would make Burns & Vollmeyer's model an interesting candidate for the modelling of insight problems, but to this thesis' author's knowledge, this has not been attempted yet.

¹⁴ Note that also for multi-space models, the requirement of a *single locus of activity* is being kept, i.e. there might be several problem spaces, but there still is only one operator being applied to one state at a time.

¹⁵ The possibility of problem-solving taking place in multiple spaces was already mentioned earlier (in passing remarks), e.g. in Newell (1967, p.28f), but Simon & Lea (1974) is usually considered the first explicit discussion of a such a model.

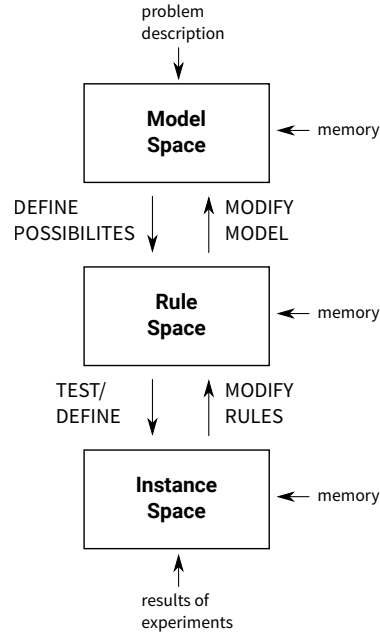


Figure 5: Burns & Vollmeyer (2000)'s 3-space model of problem solving. After Figure 1 from (ibid., p.630).

Schunn & Klahr (1995) also propose a multi-space model, namely a 4-space model aiming to explain scientific discovery. Essentially, they propose to divide each of the two spaces from Klahr and Dunbar's SDDS into two new ones: They divide SDDS' hypothesis space into a *data representation space* in which representations and data abstractions are chosen and a *hypothesis space* in which hypothesis about possible *causal relations* between the data are searched¹⁶. Similarly, the experimental space is divided into a *experimental paradigm space*, in which kinds of experiments which could inform current hypotheses are selected, and the *experimental space* proper, in which the experiments are finally carried out.

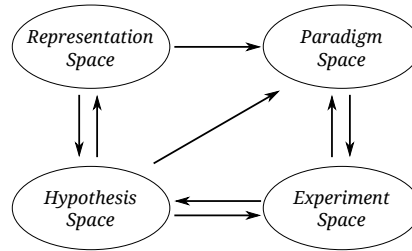


Figure 6: Schunn & Klahr (1995)'s 4-space model of scientific discovery. After Figure 1 from (ibid., p.106).

Note that all of these multi-space models are models of (some kind of) problem-solving *in general*. They are neither specific to a task domain, nor to a group of people or single person. Thus, they are more like templates for the kinds of problem space-organisations we might encounter when people

¹⁶ The distinction between the form of data and the causal relation between data, and the interaction between the two will also play a prominent role in our task, cf. Chapter 3. In cognitive science literature, very similar distinction have been known under surface similarities vs. deep structure (e.g. Chi & VanLehn, 2012) and functional vs. predicative cognitive structures (Schwank, 1999).

solve actual tasks. Thus, we cannot guarantee that for any specific protocol we can actually find all these spaces – or only those spaces. To this end, Schunn & Klahr (1996) suggest several *logical*, *empirical* and *implementational criteria* which should be fulfilled before introducing new problem spaces into a model:

- Logical criteria:
 - Spaces need to be defined such that they are unambiguously different
 - Distinction between spaces should be categorical, not gradual
 - Distinct spaces should involve distinct *goals*, and distinct *entities* being searched
 - What is “allowed” is passing of information from one space to others
- Empirical criteria:
 - Should take place at different *times*
 - Should involve different search *heuristics*
 - Behaviour should be influenced by different factors in each space
 - There has to be actual activity in every space
- Implementational criteria:
 - The spaces should be able to be represented in a computational model which can perform the task
 - Search through a specific space should be distinct from information passing between spaces

In summary, multi-space models of problem-solving try to widen the explanatory scope of classical problem space theory by assuming that *changes in one problem space are facilitated by activity in other problem spaces* with which the first one is interacting. While this allows to understand subject behaviour on a wider scope of tasks, such as inductive learning or toy models of scientific discovery, assessing a multitude of possibly changing problem spaces *empirically* also makes conducting studies much more difficult. Thus, multi-space models have actually been employed quite rarely beyond the examples quoted above (e.g. Ruiz & Newell, 1989; Qin & Simon, 1990; Kistner et al., 2014).

1.3.6 Implementational level: Problem spaces in Soar

Before we can return from this discussion of problem space notions and start refining our question, there is one further and very distinct technical use of the term problem space we have to consider, namely how it is used Newell’s cognitive architecture *Soar* (Newell, 1990; Laird, 2012).

In *Soar*, the idea that there is an ever-changing hierarchy of interacting problem spaces finds its culmination: Following Newell’s *problem space hypothesis* that problem spaces are the fundamental organisational unit of all human goal-oriented symbolic activity, *all* cognitive activity in *Soar* takes place in problem spaces (Newell, 1979). Through the process of *universal sub-goaling*, every impasse *Soar* encounters – e.g. because it lacks knowledge or does not know how to decide between alternative operators – leads to the formulation of a new sub-state and sub-goal which guides search in a new, hierar-

chically subordinate problem space (Laird et al., 1986; Laird, 2012, p.63f). In the words of Allen Newell, cognition in Soar is “problem spaces all the way down” (Newell, 1990, p.180f), with its underlying theory called the *Problem Space Computational Model* (PSCM) (Laird, 2012, chap.3).

It is important to note that the problem spaces dynamically created following an impasse in Soar’s progress on a task do not necessarily reflect the single-digit number of problem spaces we might introduce when analysing a subject’s verbal protocols according to Schunn & Klahr’s criteria above. Whereas in the latter case problem spaces are an analytical tool for problem-solving psychology¹⁷, Soar is based on the premise that goal-directed cognition is actually *implemented* in terms of problem spaces. So, continuing the analogy with Marr’s levels of analysis used above, we might call the problem spaces of Soar’s PSCM – such as task operator selection spaces, sub-task operator evaluation spaces etc. (Newell, 1990, p.181) – *implementational-level problem spaces*.

Summary

This concludes our overview of problem space theory and different uses of the concept of problem space. We introduced computational-level problem spaces which aim to be an objective description of the task environment and can be used universally to describe subject behaviour. Then we introduced representational-level spaces which include subjective notions of problem solvers thinking about a problem and hence can help explaining their behaviour¹⁸. We mentioned the possibility of these spaces changing with experience as well as due to insight. Multiple problem space models were introduced which try to explain this change by assuming it is brought about by activity in hierarchically superordinate problem spaces, but which are difficult to assess empirically. Finally, we mentioned the organisation of implementational-level problem spaces in the cognitive architecture Soar, which brings an interesting dynamic to the interaction of problem spaces, but one that does not necessarily match the psychological dimension of representational-level spaces.

1.4 MAIN QUESTION: WHERE DO PROBLEM SPACES COME FROM?

With all these possibilities it seems like we have a good framework to describe how a person is solving a problem. But even though multiple-space models seem to explain in part how a problem space is constituted and there is some research on how problem spaces are chosen initially (cf. p.9), there is still no fully satisfactory answer to the question (cf. also Ohlsson, 2012b)

Where do the problem spaces come from? Or better: How are they constructed? Why are their constituents chosen as they are, why and

¹⁷ As they were in most of Newell & Simon’s earlier work, cf. for instance Newell (1967, p.49): “The problem[-behavior] graph is a projection of the total behavior of the subject into a space of our own devising.”

¹⁸ Note however that there is no standard terminology to distinguish between these cases. For instance, while Newell & Simon (1972) use the term ‘problem spaces’ to varyingly refer to computational-level spaces (sometimes with the qualifier “legal”) and representational-level spaces pertaining to individual subjects, Reed (2015) calls the former ‘problem space’ and the latter ‘search space’.

when do they change, and what mechanism mediates this set-up and subsequent changes?

Leaving aside Soar's psychologically unsatisfying answer¹⁹ and even though Newell & Simon acknowledged this unsatisfactory state of affairs early on (Simon & Newell, 1971, p.154), it has been noted that problem space theory never managed to provide a convincing answer (Ohlsson, 2012b, p.118).

Acknowledging the importance of problem space set-up, VanLehn introduced the notion that problem solving is "*understanding + search*"²⁰, implying that problem space search can only commence once we have understood a problem, and also mentioning that most likely, the two processes are *interleaved* (VanLehn, 1988, p.5f; cf. also Simon & Hayes, 1976). Proceeding from this distinction we could say that problem space theory is very informative with regards to search, but does not tell us much about understanding (in VanLehn's sense). Therefore, this thesis should aim to elucidate what happens when people try to understand a problem, in concert with trying to solve it.

1.4.1 *Reasons for the de-emphasis of problem understanding*

What could be the reasons that problem solving research still lacks knowledge about problem understanding? VanLehn (1991) suggests that this is due to a lack of evidence and informative data, mentioning that existing data can be modelled equally well by a large number of different approaches (ibid, p.2).

In this thesis, we claim that one reason why existing data is not sufficiently informative for theory formation is because research focussed on so called *well-defined problems* (WDPs) (Minsky, 1961, p.9). According to Minsky, WDPs are those for which there is a systematic procedure for deciding what constitutes an acceptable solution. This entails a "canonical" representation for problem states and a closed set of permissible operators. Even if we are using different cover stories for our problem, or other forms of problem isomorphs, for a WDP there is always a more or less straight-forward mapping from the task instructions to a problem space in which a solution can be found (Hayes & Simon, 1974; Simon & Hayes, 1976), in other words: There is comparatively little "understanding" to be done.

Based on our distinctions introduced earlier, we could say that a problem is well-defined if its computational-level problem space and the representational-level problem spaces most likely to be chosen initially by subjects are either coextensive or at least very similar. And if we concede that many problems people actually face in real life are not WDPs, then using them in research on how problem spaces are set-up will necessarily fall short of giving us a complete answer.

¹⁹ Soar's axiomatic answer to this question is that we *always* are in some problem space, be it a global top-level space or some subordinate space, with new sub-spaces being constantly built until the process bottoms out when every piece of knowledge needed to act is available. So ultimately each problem space in Soar is either "innate" or created by universal sub-goaling.

²⁰ Note that what VanLehn calls "problem understanding" is very similar to what Gestalt psychologists called *problem perception*, emphasising that the initial understanding of a problem can often depend on the perceptual processing of its presentation (Köhler, 1947).

The opposite term – *ill-defined problems* (IDPs) – was introduced by Reitman (1965)²¹. The property of ill-definedness is usually defined ex negativo as the absence of well-definedness: Any problem which does not meet Minsky's criterion should be considered ill-defined. From our earlier remarks about the subjective nature of problem spaces it should be apparent that this distinction depends highly on the problem-solver and is not dichotomous but rather gradual. Depending on how strictly we approach a problem and what constraints we demand a solution to fulfil, a problem might appear equally well- or ill-defined. In fact, Reitman already discussed the continuity between the two notions himself:

“(...) to the extent that a problem situation evokes a high level of agreement of a specified community of problem solvers regarding the referents of the attributes in which it is given, the operations that are permitted, and the consequences of those operations, it may be termed unambiguous or well defined with respect to that community.” (Reitman, 1965, p.151)

Simon (1973) criticises this negative definition, remarking that it is very difficult to prove for any negatively-defined category that it is empty. He proposes that even when solving an IDP, a problem solver always works in some – if maybe inadequate – well-defined problem space, and through an *additional process* of ongoing recognition and retrieval of possible operators and alternative state representations this problem space is being modified until it is sufficiently adequate to solve the problem²². It is important to note that even with this eliminative explanation of IDPs by Simon, the understanding process appears to be much more involved in the problem solving process than it would need to be for a WDP (cf. also Reed, 2015).

There are also researchers with dissenting opinions, who claim that the process by which we understand a problem (conceptualised as the transformation of an IDP into a WDP) relies on fundamentally different mechanisms (e.g. Goel, 1995, 2010). It is an open, if maybe metaphysical question whether there are problems which are ill-defined *in principle*.

Independent of the eventual nature of IDPs however, the fact remains that most research in problem solving, especially studies employing problem space theory, is focussed solely on the study of WDPs. As the short discussion of IDPs above was supposed to demonstrate, this focus leads (a) to an underestimation of the difficulty of *understanding* a problem, and (b) to the assumption that we can neatly separate the study of problem-solving (search) and understanding. Therefore, we still lack research enabling us to answer the question where problem spaces come from (how they are set-up, how they change etc.).

The phenomena of problem space set-up, change etc. are what we comprehensively called *representational dynamics* in the thesis title. Since “representational change” is already used in a somewhat narrower sense, especially in research on insight, and since the term “change” shrouds the breadth of the phenomena – from gradual transformations over changes of levels of

²¹ Although most researchers use the terms well- and ill-defined, Newell, Simon and some of their students used the practically synonymous terms well- and ill-*structured*, first in Simon & Newell (1958).

²² This scheme of working in a problem space which is modified by a hierarchically superordinate process was first proposed in Newell et al. (1959), which subsequently developed into the dual-space model proposed in Simon & Lea (1974, see discussion above), and ultimately lead to Soar's PSCM.

abstraction to switches between cross-domain descriptions – we settled for the broader term “dynamics”²³.

1.4.2 *The methodological dilemma*

It is important to note that this lack of knowledge is not arbitrary. In fact, there is a solid methodological reason for focussing on search and WDPs: Remember our paraphrase of WDPs as problems whose computational and representational-level problem spaces are likely to be very similar (cf. page 17).

This means that for a WDP we can (to some extent) predict the space in which problem solvers will work on the problem. This in turn allows us to work in a hypothesis-testing experimental setting, where we can compare results as well as performance of our subjects. If we can even predict (or hypothesise) that the choice of (one of a small number of possible) problem spaces depends on a specific variable and has specific effects on progress and outcome of our problem solvers, we can conduct sound comparative, experimental studies.

If however, we cannot – or only to a much smaller extent – predict how the problem space a subject will employ looks like, we also cannot make precise predictions about the progress or results of them working in these unknown spaces. And ultimately, *if we cannot ensure that people will work in the same, or very similar, problem spaces we cannot easily compare their performances.*

So, for the time being, when investigating IDPs we either have to assume a coarser level of analysis and develop a coarser theory (such as assuming a computational-level perspective), or we have to loosen the methodological constraints we are setting us, and proceed to work not in a hypothesis-testing manner, but in a theory formation mode, and to *conduct exploratory case studies.*

After considering existing approaches towards the investigation of IDPs, in the following chapter we will argue for working in the latter mode and assess the properties such a case study ideally should have.

1.5 SUMMARY

This concludes the introductory chapter. We started out with introducing human problem solving as the main concern of this thesis. We introduced the classic theory of problem solving, based on Newell & Simon’s work, most importantly the concept of problem spaces. We distinguished three different uses of the term problem space, and likened those uses to the three different levels of analysis – the computational, representational, and implementational level – introduced into cognitive science by David Marr (Marr & Poggio, 1976).

We noted that there are phenomena in human problem solving – most notably learning and insight – which require us to assume that problem spaces are not static entities, but that they can change both suddenly and evolve

²³ The breadth of this umbrella term also means that it should not be confused with a technical term. In most parts of the thesis more specific terms for the phenomena described will be used and occurrences of the general term might be quite rare.

over time. We presented several models of hierarchically interacting, multiple problem spaces which are supposed to explain how such change is facilitated. We noted however, that only few studies have been conducted with such multiple problem space models, suspecting this might be due to the difficulty of using them as analytical devices in empirical studies. Our treatment of problem spaces left us with the assessment that classical problem solving theory still cannot explain how representations are set-up and how they change, summarising these phenomena under the umbrella term *representational dynamics*.

We proposed this lack was due to most of research on problem solving concentrating on well-defined problems (WDPs), attributing this concentration to a methodological reason, namely that the more WDP-like a problem is, the easier it is to predict people's representational-level problem spaces which in turn enables experimental studies and between-subject experimental designs. We recommended this lack of theory might be faced with an investigation of ill-defined problems (IDPs) and a loosening of methodological constraints, proposing to conduct *exploratory case studies*.

1.6 OVERVIEW OF THE REMAINING THESIS

After this introduction to problem space theory, and having identified the need for an investigation of representational dynamics and ill-defined problem solving, in the remaining thesis we will discuss a study we have designed with this interest in mind. The main two deliverables of this thesis will therefore be (a) a description of this study's design and method, and (b) a detailed analysis of the task we have designed to be used in the study (iterated, mental paper-folding).

In detail,

- Chapter 2 will discuss design, setup and methods of the proposed study
- Chapter 3 will present our task and the current state of our cognitive task analysis, based on an analysis of formal domain properties, insights from pilot studies, and present tentative problem space notions for dealing with these insights
- Chapter 4 will conclude the thesis with a general discussion of our study in the context of problem-solving research, and proposed avenues for future work.
- Appendix A contains the study material, such as consent forms and task instructions
- Appendix B will present the detailed guidelines investigators are supposed to followed in our study
- Appendix C contains and describes the software written for the thesis

DESIGN AND METHOD

"(...) what seems to be missing is an attractive problem solving theory that stimulates further research, a proper set of diverse tasks for experimenting with them, and a methodology that opens again the window to inner processes."

— Funke 2014, p.496

So how could we investigate IDPs and representational dynamics? In this chapter we present the design, method and setup of a study that we propose could approach these questions. Since the study design as well as our chosen task deviate in several ways from standard procedures in problem solving research, we will first discuss the key properties of our design rationale, before describing the concrete study.

This discussion starts with a look at existing research paradigms with respect to IDPs and representational dynamics in contrast to which we derive the properties of a task suitable for our purposes (Section 2.1). Then we continue with a discussion of some of the key properties of our study design (Section 2.2), and conclude with an overview of the properties of both, task and study design (Section 2.3).

The description of the concrete study starts with a short overview of the study (Section 2.4), and then discusses the details about the study's tasks (Section 2.5), setup (Section 2.9), procedure (Section 2.8), preparation of subjects and investigators (Section 2.6), and our approach towards analysis of the gathered data (Section 2.10).

The next chapter will then proceed to discuss a detailed analysis of the iterated mental paper-folding task we present here, developed from data gathered in pilot runs of the study presented here, as well as from formal considerations.

2.1 DESIGN: TASK PROPERTIES

In this section, we start out by looking at existing branches of problem solving research (beyond problem space theory) which try to find out how people solve problems which require more change than classic tasks (i.e. which are – to some extent – IDPs). While we do not directly adopt either approach for our own study, we will introduce them briefly and introduce our choices of task properties in contrast to them.

2.1.1 Existing approaches: Insight problem solving

The field of insight problem solving was already examined shortly in the introductory chapter (cf. p.11). Chu & MacGregor (2011) provide a current

review of the most important empirical results. Since insight problems¹ allow people to represent them in multiple manners (namely those in which they are able to see the solution and those where the solution is impossible to achieve), they are ambiguously termed and hence could be described as IDPs. This should make them good candidates for investigating change of representation and understanding.

Yet, insight problems are a very special kind of IDP, which we might call “deceptively well-defined”: The problem can be stated concisely and hence subjects usually employ relatively predictable problem spaces to solve them, only (according to Ohlsson’s Representational Change Theory) one which is overly constrained in some manner and precludes the subject from finding the solution (cf. Figure 3). After their original problem space has been adapted, subjects can work towards the solution in a new, equally predictable problem space. While this calls for *some* representational change, the problem spaces involved are still predictable enough to allow research in standard experimental settings, striking an interesting balance between ill- and well-definedness.

The main problem with using insight problems in investigating representational change is that for most insight problems this change appears to be *singular* and *sudden*². This puts a sharp limit on the amount of empirical data we can gather about this change. Researchers are not certain exactly which components of insight are brought forth by conscious or unconscious processing, but subjects’ reports certainly contain events which appear “out of the blue” and could thus be hypothesised to be the outcome of an unconscious process (Chu & MacGregor, 2011, p.143f).

Therefore, choosing an insight problem as task for our study might not be optimal. In order to increase the chance that our data might include evidence of representational dynamics, we should instead try to find a task which is less well-defined and affords more change. Hence, we propose to use an IDP-like task which gives subjects *more freedom in the construction of their problem spaces*.

This could for instance be done, by allowing subjects to *find novel ways of representing the problem*, and by using a task which involves *multiple changes over a longer stretch of time*. As we discussed in the introduction, such more gradual changes take can place through learning, thus our task instruction should aim at the *facilitation of learning*.

2.1.2 Existing approaches: Complex problem solving

Another branch of problem solving research which grew out of dissatisfaction with the simplicity of classical problems described above works on what they have termed “complex problem solving” (CPS) (Frensch & Funke, 2014)³. Aiming for greater *ecological validity* in the choice and scope of tasks,

¹ Note that not every person working on an “insight problem” does necessarily experience an affective Aha-moment when solving it (Ohlsson, 2011, p.90), and in turn insight might be experienced without engaging in problem solving (Chu & MacGregor, 2011, p.120). Hence there is no strict definition of what constitutes an insight problem, except for saying that a problem is an insight problem when most people solving them report a feeling of insight.

² The perceived suddenness of its origin is usually considered the defining property of insight (cf. Poincaré, 1913; Van der Waerden, 1953; Hadamard, 1954).

³ More recently the research group around Funke started to call their field “Dynamic Decision-Making” instead (cf. Fischer et al., 2015).

Dörner & Kreuzig (1983) define a problem as complex once it fulfils the following criteria (as cited in Funke, 2010, p.134):

- Its constituting elements are large (*Complexity*)
- They are highly interconnected (*Connectivity*)
- They change over time (*Dynamics*)
- Their structure and their dynamics are not given explicitly (*Intransparency*)
- There are several facets to the goal which have to be coordinated (*Polytely*)

Typically, their tasks are implemented as so-called *microworlds*, computer simulations of complex scenarios fulfilling the above criteria, in which subjects have to try to successfully act for a given amount of time. Well-known examples for CPS microworlds include the LOHHAUSEN scenario (Dörner et al., 1983) – in which subjects have to manage around 2000 interacting variables as the mayor of a small town (Goal: “Take care of the well-being of your town’s population!”), and the TAILORSHOP scenario – in which subjects have to manage a small clothes production facility (Funke, 2003, p.146).

Due to the complexity of these scenarios, they are not so much solved as being “handled” by subjects⁴, and it is apparent that subjects represent them in very different, individual problem spaces (Kluwe, 1995, p.270). Hence, CPS microworlds definitely constitute IDPs according to the definitions above.

Owing to this individuality, most experimental studies of CPS concentrate not on the representations employed by subjects, but on an analysis of the influence of cognitive and situational variables on CPS performance. These variables include e.g. general intelligence, expertise, motivation and affectively relevant manipulations of the task environment (Wenke & Frensch, 2003; Barth & Funke, 2010). Typical measures beyond task performance include broader classifications of behaviour such as the number of actions taken in a certain amount time, or the percentage of time spent with the gathering of information before taking action.

Using these and similar measures, CPS researchers have gathered many empirical correlations and started to formulate a process theory of CPS (Fischer et al., 2012)⁵. Since problem representations of CPS tasks are likely to be very idiosyncratic and thus cannot be easily assessed through the analysis of averaged data, studying them would require an investigation by the means of *single-case studies*. While there have been repeated calls for conducting case studies, to the knowledge of this thesis’ author there have only been very few (Kluwe, 1995; Funke, 2003). Therefore, since CPS research usually does not aim to elucidate the individual problem spaces used by their subjects, the current state of their theory cannot tell us a lot about how they are constructed.

The take away from CPS research for our question seems that when we are interested in situations where subjects constructing individual, changing problem spaces, we should use tasks with high complexity and vague goals (vague as in incompletely specified, not as in nebulous) (cf. also Burns & Vollmeyer, 2002). However, for our purposes CPS tasks overshoot the mark somewhat, by essentially being designed in a way that a subject has virtually

⁴ Certainly a feature of many real-life problems, granting the CPS approach high ecological validity.

⁵ Yet, in larger parts, this theory draws on classical problem solving theory, whose applicability to CPS remains an open question (cf. Quesada et al., 2005)

no chance of grasping their underlying structure. Hence, even though the problems themselves are very complex, it might actually be that the problem spaces a subject uses for dealing with such a problem are not complex in themselves – bringing forth relatively simple behaviour such as the VOTAT strategy (“Vary one thing at a time”) (Tschirgi, 1980; Fischer et al., 2012, p.31).

But regardless of whether a subject’s representations of a complex problem are complex in themselves, using a problem involving as many (partly hidden) variables as CPS tasks usually do, it would be very difficult to infer which variables play a role in a subject’s problem space and which do not. Ideally, we would need a task which seems complex at the outset, but which still is *understandable in principle*, or in other words *a task which confronts subjects with high complexity at first, but which subsequently allows this complexity to be reduced* up to a point where it is conceivable that a subject could represent a current problem state in their head.

2.1.3 Task properties

Summarising what we take away from insight and complex problem solving research, we would like a task with the following properties:

- (1) Solving it requires *multiple changes* in problem representations
- (2) These changes should take place over a comparatively *long stretch of time* (granting more opportunities for observation),
- (3) The *initial goal should be vague*, requiring the subject to work out by themselves what they should do
- (4) The *complexity* should be *high, but reducible* over time, by finding other representations

Another way of saying that a task requires the subject to make changes to their representations (Point 1) is that it *requires learning*. This means, we should ensure that such changes are not only possible, but that they actually *occur*. This also means we should take care not to pick a task for which people might have preferred representations with which they might stick for the duration of the whole study.

Point 2, enlarging the time frame of the study – might go some way of increasing the likelihood of changes of representation, but especially an initially vague goal (Point 3) and a level of complexity which “overstrains” naïve approaches to representation (Point 4) should ensure that every subject working on such a task will over time employ several different representational approaches.

2.2 DESIGN: KEY PROPERTIES OF THE STUDY

Conducting an empirical study of a task with the properties sketched above has also important implications for study design and methodology. Some of the properties discussed below have already been mentioned in passing in the introduction, but here each of them is explicitly discussed in a brief manner.

Note that while some of these properties would merit a far more extensive methodological and philosophical discussion, since this thesis' focus lies on the theory of problem solving and a discussion of our paper-folding task in the context of this theory, we can only touch lightly on the issues this might raise.

2.2.1 *Exploratory single-case studies*

Empirical studies can be conducted either with the aim of *hypothesis testing* or *theory formation* (de Groot, 1965, chap.9). After the early work by Newell & Simon, which mostly comprised exploratory studies aimed at theory formation, studies in problem solving increasingly took the form of hypothesis testing experiments (cf. Ohlsson, 2010). Since the formulation of hypotheses is ideally based on theory, the historical shift from theory-forming to hypothesis-testing studies reflected the progress made in problem solving theory. Additionally, this shift was also furthered by Ericsson & Simon (1993), who laid the methodological foundation for conducting quantitative experimental studies which use verbal reports as data.

But if experimental studies should be carried out with hypotheses derived from theory, we can only study those phenomena for which there already are solid theoretical ideas. And as the introductory chapter was supposed to show, theoretical ideas about the representational dynamics in problem solving are exactly what problem space theory is lacking. Hence, it should be clear that – for the time being – we have to proceed working in an exploratory manner, with the aim of theory formation.

Furthermore, the idiosyncratic nature of the problem spaces subjects construct for working on IDPs necessarily precludes us from grouping subjects and averaging data in a straight-forward manner. Focussing on *single cases* instead allows a detailed analysis of each subject's (idiosyncratic) representations, whereas a comparative study would require us to abstract away from many of these particularities until a comparable common ground between subjects had been identified (cf. VanLehn, 1991; Kluwe, 1995; Funke, 2014). Yet, without prior knowledge about the subjects' problem spaces, we have to assume that all idiosyncratic behaviour might eventually turn out to be functional and relevant to the problem's solution, and we should attend to it.

While it may still turn out *afterwards* that two subjects used very similar mental structures and that their data might therefore be comparable, we do not know beforehand how similar or dissimilar two subjects' problem spaces will be, hence in the general case we cannot assume that they are comparable. Only *after analysis* of their protocols, we can ascertain (dis-)similarities between the cases.

We should however keep in mind that if we generalise over these similarities – contrary to when working in a hypothesis-testing manner – these generalisations do not state something about the population, but are theoretic conjectures. This important caveat should not deter us from making such generalisations altogether, since they form the basis for theory formation (Kluwe, 1995; Schoenfeld et al., 1993).

2.2.2 Microgenetic studies

Our study proposes that in order to understand better what problem spaces are, we have to investigate their change, growth and general dynamics *over time*. The branch of psychology generally concerned with how cognitive structures, capabilities etc. change over time is *developmental psychology*, and the view, that phenomena are best understood through their history, is widely held in developmental psychology and the central tenet of Piaget's *genetic epistemology*⁶ (Voneche, 1999).

Yet, developmental psychology usually deals with far larger time frames of development (months, years, even decades, cf. Granott & Parziale, 2002), and we should not expect that any of their study designs would be directly applicable to problem-solving and learning, which takes place over minutes, hours or days. Yet, in contrast to this more usual "macro-perspective", there are developmental and educational psychologists who look at such time spans and conduct what they call *microgenetic studies*⁷ (Siegler & Crowley, 1991).

According to Siegler (2006, p.469) there are three defining properties of a microgenetic study:

- (1) Observations span a period of rapid change,
- (2) Within this period, the density of observation is high (relative to rate of change), and
- (3) Observations are analysed intensively, with the aim of inferring underlying representations and processes

Typically, such studies comprise several trials of working on the same, or very similar tasks, with data being gathered either continuously, or in between single trials (e.g. with retrospective interviews). This data is then analysed in a trial-by-trial manner, looking to identify how the subject's behaviour changed over the course of the trials (Siegler, 2002, 2006).

One of the most prominent observations made in microgenetic research is that there appears to be much more variety in people's behaviour than most theories of cognitive development and skill acquisition can explain at the moment. The tendency that people might e.g. sometimes use an older strategy towards a task even though they had already learned a newer, better strategy, lead to what Siegler called the *overlapping waves theory*: Cognitive strategies do not supersede each other in an orderly, discrete fashion, but the use of different strategies might overlap, with their relative frequencies depending on many factors (Siegler, 2006, p.477f).

For our study, we propose to *use a microgenetic setup*, which has implications for our task, our study design, and our approaches for data analysis. Our task should involve change and consist of sub-tasks stretching over several trials (sessions), and our methods of data gathering and analysis should allow gathering data about this change and making inferences about it respectively. While most microgenetic studies are not case studies and do not focus on representational dynamics, none of their defining principles conflict with this focus (Fazio & Siegler, 2013). For instance, Parnafes & diSessa

⁶ The term "genetic" does not pertain to the modern biological sense as in genetic material, but to "genesis", i.e. the formation or development of something.

⁷ Sometimes this approach is also called *microdevelopmental* (cf. Granott & Parziale, 2002). For a short history of the concept and its names, see Karmiloff-Smith (2013, p.47f).

(2013) proposed a microgenetic setup for qualitative analysis of conceptual change and learning in a tutoring situation, which was also a large inspiration for our approach towards data analysis (cf. Section 2.10).

2.2.3 *Acquiring process-level data*

There are a multitude of empirical data sources employed in problem solving research. Roughly they can be categorised as either *performance measures* or *process measures*. Performance measures such as reaction times or task-specific scores measure an outcome of a subject solving (or not solving) a task or circumstances of this outcome. With specific hypotheses, such measures can be efficiently used to e.g. compare the relative performance of two different solution strategies for the same task. But here, once more, we face the issue of our necessarily limited prior knowledge of how a subject will solve an IDP. This in turn limits our ability to formulate precise hypotheses, making performance measures become less informative for our case.

More informative, but also harder to acquire and analyse are measures, which are gathered continuously throughout a subject's engagement in a problem-solving process⁸. These include the recording of behaviour in the task environment, eye tracking, EEG, different forms of audio/video recordings and verbal reports. While eye tracking and EEG are gathered continuously and provide data which can be analysed exploratorily, a meaningful interpretation of these data still relies on the researchers having some knowledge beforehand about what the subject is actually going to do (which we do not, for IDPs). Thus, they are less well suited for a broad, exploratory investigation of IDPs such as ours.

Verbal reports and introspection

Depending on the chosen problem, there might be lengthy episodes in problem solving where a subject shows no observable, external behaviour and instead just "sits there thinking", i.e. recording the subject's observable behaviour might turn out to be not enough to infer to a subject's internal representations on its own. Therefore, in many problem solving studies further process measures are begin gathered, such as video recordings and verbal reports⁹. Most typically, researchers instruct subjects to use report on their thinking using a special form of concurrent verbalisation known as "Think aloud" (TA), record these verbalisations, and analyse these *protocols* (de Groot, 1965; Ericsson & Simon, 1993). But there are also other kinds of verbal reports beyond TA, such as systematic introspection, explicitation interviews, and Socratic dialogues (Kölpe, 1920; Petitmengin, 2006; Stenning & van Lambalgen, 2004).

One problem with gathering (any kind of) verbal reports, is the possibility that "verbal overshadowing" might occur (Schooler & Engstler-Schooler,

⁸ Note that none of these "process measures" actually is a *direct* measure of the internal problem-solving processes. In fact, they are merely measures which lend themselves to quasi-continuous, high-resolution assessment, and hence are suitable to indirectly observe a process unfolding over time.

⁹ Which one of these makes sense, again, depends on the chosen problem. E.g. if we expect that gestures, movement, facial expressions etc. could play an important role in our task, then we might prefer video recordings.

1990). Overshadowing is an interfering effect that might occur when subjects have to verbally describe non-verbal thoughts which is hard to put in words¹⁰. Therefore, in our study, we try to alleviate this effect by offering subjects the alternative *option to draw sketches* of observation they feel would be difficult to express verbally.

The main data acquisition method employed in our study is a mixture of *nearly-concurrent* (“event-contingent”) *introspection and retrospective interviews*. Since the use of introspection is usually met with a lot of (partially warranted) scepticism in experimental psychology (cf. Nisbett & Wilson, 1977; Mandler, 2011), we will briefly argue why we think that a carefully conducted and restricted form of introspection could be useful.

We assume that the use of introspection for data acquisition gives us access to information about the subject’s thinking beyond that which a think aloud-methodology could yield. This assumption has its origin in research conducted by a working group surrounding Oswald Wiener (cf. Raab & Eder, 2015). The specific argument we refer to here is made in detail in Jäkel & Schreiber (2013). Mainly, it is based on the observation that under specific circumstances – namely when solving problems which necessarily involve learning – introspection could be to some extent considered a natural part of the problem-solving process. This also resonates with a large amount of anecdotal evidence about the importance of introspection especially in mathematical problem solving (Hadamard, 1954; Poincaré, 1913; Pólya, 1945). This would mean that if we were to take special care in instructing subjects in such a way that they do not introspect (as is done in TA setups, cf. Ericsson & Fox, 2011) this would not only prevent us from observing these episodes, but it would prevent them from occurring at all, hence removing a natural ingredient of problem-solving behaviour (Jäkel & Schreiber, 2013)¹¹.

Of course we are aware of the scepticism brought forth against using introspection as a method in psychology (mainly referring to Nisbett & Wilson, 1977), but it is our impression that this criticism has been exaggerated (cf. also Overgaard et al., 2008; Raab, 2015a). Nisbett & Wilson (ibid.) review the results of several studies, concluding that introspective self-reports are not a reliable measure, however the studies they quote have in common that subjects were asked for the reasons of *why* they have done something (as opposed to reports of *what* someone has been doing), and they were asked *retrospectively*. While this is an important indication that we should take introspective reports about *causal relations* of thoughts with a grain of salt – especially if they are given in hindsight – it is no reason to discourage the use of introspection altogether (cf. also Siegler & Stern, 1998). While a recent meta-analysis has shown that concurrent verbal reports can be a reliable measure (Fox et al., 2011), Siegler & Chen (1998) indicate that *immediate retrospection* (in a 2–30s time frame) can also be used reliably.

¹⁰ In some cases this effect can even be so severe that it makes it impossible to retrieve a visual memory after having to verbally describe it (Schooler & Engstler-Schooler, 1990, p.41).

¹¹ Thus, this should be a testable hypothesis: If introspection facilitates learning in complex tasks, subjects verbalising in a TA fashion should have fewer learning events (and consequently lower performance on transfer tasks) than those who were instructed to introspect. The main difficulty would be inter-subject comparability if we were to use the kind of complex IDPs we want to investigate.

Training and instruction

Apart from restricting the time frame of reporting and taking care that the subject will not engage in speculative causal reasoning, important measures we take to ensure data reliability is that our study include an extensive *training session* and carefully designed *instructions*.

Since our study design requires the investigator to be very well acquainted with solving the tasks themselves, they have at the same time to be very careful to not impart specific task information, opinions, or theoretical convictions to the subjects. By the same token, we have to assume that subjects have their own folk-psychological notions and attitudes, which are similarly subject to any number of biases.

Hence, the training session cannot consist of the investigators explaining the subject “how it is done”, and neither should the subject simply rely on their own predilections. Instead, the investigators should in a neutral manner *aid the subjects in making their own experiences* with problem solving, introspection and self-reporting.

The *instructions* the subjects receive are designed to make them aware of possible error sources (such as omissions, commission, verbal overshadowing, speculative reasoning, ...) – once more – without telling them what to do. We describe the instructions in Section 2.7 and the details of the training session in Section 2.8.1.

Written reports

One peculiarity of our study design is that we ask our subjects to *write* their self-reports instead of giving them orally. While this may sound methodologically questionable, there are several reasons which motivated this setup¹²:

(1) Having the subject write their report allows the investigator to *read along* (on a computer screen of their own) and think about the subject’s report in *real-time*. This allows them to make notes about unclear or *ambiguous episodes*, which they can inquire about in a *retrospective interview* taking place immediately after the subject has finished working on the task. (2) while writing might interfere more strongly with the task than talking, our setup already accepted that our method of data acquisition is *reactive* with respect to our task. Hence, the exact amount of interference is only a secondary concern. (3) with our setup of intermittent reporting (which usually leads to patterns of thinking for 30 seconds to a few minutes, and then writing for about a minute), allowing the subjects to write feels like a smaller interruption of their thinking, since they do not have to “break the silence”. (4) we assume that given the ubiquity of written communication via the internet, writing colloquially in a chat-like interface actually feels more natural to the young students who usually are our subjects than it might have for older generations. And finally (5) we assume that written language, more so than spoken language, makes the subjects express themselves *more precisely* (e.g. using fewer pragmatic or gesturing references instead of naming the referents).

¹² Note however that due to limited resources on our part, these reasons have yet been evaluated empirically (nor do we know of existing relevant evidence).

Self-study of the investigator

There is a further property of our study design concerning introspection, which is based on the observation that when analysing another person's reports, the *investigator* has to some extent to rely on their own introspections. Of course their interpretation should be derived from reasonable inference, checked with literature and other researcher's interpretations in order to minimise bias, but at some point the investigator cannot escape trying to understand someone else's thinking by starting out from their own, basing their interpretation on "empathetic guesses" of what happens in the other mind (Raab, 2015a, p.25) – at least as long as there is no fully-fledged formal theory of mental events which could be used to generate an interpretation instead.

Therefore, and also given the complexity of our task and methods, it is important that the investigator is extremely well prepared when conducting the study. In order to quickly recognise typical errors, to be able to understand the subject's protocol and to guide the interviews, the investigator needs to be experienced in (a) introspecting themselves, (b) problem-solving in mental imagery, (c) in solving the paper-folding task themselves, and finally (d) they need to be knowledgeable about possibilities of thinking about the paper-folding task beyond their own idiosyncratic approach.

Thus, instead of leaving the investigator to make these observations privately and undocumented, our study requires the investigator to acknowledge their own thinking by *conducting a self-study* before proceeding to conduct the study with someone else. This way, the specific perspective the investigator necessarily has towards a subject's data is, at least to some degree, documented.

2.2.4 *Mental imagery*

Another key properties of our approach is to conduct studies of problem solving *in mental imagery*, i.e. asking the subjects to work on a task purely mentally and without external aids. There are several reasons why we think this is crucial:

First, asking subjects to work in imagery has been shown to *facilitate learning*, which Sweller and colleagues termed the *imagination effect* (Leahy & Sweller, 2008; Sweller et al., 2011). Sweller et al's hypothesis is that by having to work through an entire task solely in working memory, the probability is increased that information is transferred to long-term memory. Related effects are also described by researchers outside of Sweller's group, variably known as *mental practice* (Clark, 1960), or *mental rehearsal* (Dunbar, 2000).

Additionally to facilitating learning, working in imagery is associated with *creativity*. Finke (1993) describes several studies which showed the beneficial effects of mental imagery on creative concept invention. Similarly, Reed (1993) shows that using mental imagery can be an efficient strategy for mental synthesis and problem solving, indicating that imagery might be used in dealing with unknown tasks, and to generate *novel representations of a problem* (cf. also Kaufmann, 1990).

Another important point of requiring subjects to use imagery, is that when we want to investigate how they construct and change their mental repre-

sentations we have to make sure that they actually mentally represent the problem. Research on situated cognition has been emphasising for a while now, that in real-world situations, people often opportunistically employ parts of their external environment in place of relying on detailed, internal representations (Hutchins, 1995; Clark & Chalmers, 1998; Kirsh, 2009a).

While we are aware that having subjects sit in a barren university room grants limited ecological validity to our setup if compared to situated approaches (cf. Dunbar & Blanchette, 2001), we have to keep up this artificiality if it is exactly these detailed, internal representation which we want to learn more about. After all, even if they not always do so, humans are able to construct internal representations, and doing so can be beneficial: Delaney et al. (2004) show for instance, that when subjects have to mentally plan their problem solving their solutions are more efficient than subjects which were allowed to directly interact with the task environment¹³. Similarly, Noyes & Garland (2003) report that subjects who could directly manipulate the task environment were more likely to engage in trial-and-error behaviour, whereas subjects who had to proceed mentally were most efficient, however at the expense of needing more time and having a greater probability of failure.

Finally, there is a methodological reason for working with imagery, which is based on an assumption of ours:

When interacting with our environment we employ sensorimotor processes, many of which are highly automatised and work without us explicitly attending to them, or even if they require attention, they might have effects of which we are at least partially unaware. This suggests, that when we attend to our perception of the environment, we are more likely to process some part of it tacitly – with it possibly having an influence on our thoughts without us knowing – than if we were to attend solely to our mental imagery.

This is not to say that there we are necessarily aware of all processes involved in mental problem solving – just that due to the limited capacity of our thinking without any sort of external aid, we might expect fewer influences from processes we were unaware of than if we were to attend to our perception and action. Which in turn could pose problems to a study working with self-reports: Maier (1931) for instance – widely quoted as evidence for the unreliability of self-reports – showed that $\frac{2}{3}$ of his subjects did not notice a perceptual hint to a problem solution, even though they could be shown to have processed it, some of them instead giving (unknowingly) confabulated reports about how they conceived of the solution¹⁴. Given that mental imagery seems to involve a gradually smaller amount of automated processing than perception, we assume that cases like the one reported by Maier should be rarer when working in mental imagery.

¹³ Crucially, this improved solution efficiency remained even when not having to work in imagery any longer, hinting once more at learning through the imagination effect. If we just look at immediate efficiency and not at lasting effects, then there are also cases (depending on the problem) where interaction with external representations improves efficiency instead of deteriorating it (e.g. Kotovsky et al., 1985).

¹⁴ It should be noted that $\frac{1}{3}$ of the subjects *did* notice the hint and correctly identified it as the source of their insight. We mention this, because Maier (1931) is often superficially cited as having shown that people's causal retrospections are confabulated *in principle*.

2.3 DESIGN: OVERVIEW OF THE KEY PROPERTIES

This concludes our explanation of the design of our task and study. The main properties were:

Concerning the task:

- (1) Task requires *multiple changes* of representation
- (2) Changes should take place over a *longer time span*
- (3) The initial *goal should be vague*, allowing subjects to find their own specifications
- (4) Task *complexity* should be *high, but reducible* by finding better representations
- (5) Task should consist of several *sub-tasks*, divisible over *multiple sessions*

And concerning the study design:

- (6) Study is *exploratory*, aiming for *theory formation*
- (7) Takes form of *single-case studies*
- (8) Follows principles of *microgenetic research*:
 - a) Observations span a period of rapid change
 - b) High density of observation (relative to rate of change)
 - c) Intensive analysis of data, aiming to infer representations and processes
- (9) Acquires *process-level data*:
 - a) Combine event-contingent, immediate retrospection and interviews
 - b) The subject should give their report in written form
 - c) Avoid verbal overshadowing by allowing sketches
 - d) Training session allows subject to gather experience
 - e) Investigator should conduct self-study
- (10) Have subjects work in *mental imagery*

In the following, we will describe how the concrete study we are proposing (and of which we conducted several pilots) is set up, fulfils properties (6)–(11), and what measures we take to deal with the potential difficulties of introspective methods. For your guidance, we start this description with an overview of the general setup and procedure, before we proceed to describe the details of the study.

2.4 METHOD: OVERVIEW OF THE STUDY

In total, there are *5 sessions*, one training session and 4 main sessions. The sessions are conducted on consecutive days and each one lasts about 2 hours.

Each session consists of two task blocks, and apart from the first block of the training session, the *procedure* for every block is similar: The subject works on their task and intermittently writes their self-report while the investigator reads along on a separate computer screen and makes notes about segments they do not understand. After the subject has finished their task, there is a

short pause after which the second, similarly structured task block commences. At the end of this second block, there is a retrospective debriefing interview in which the investigator uses their notes to guide the interview and get the subject to elaborate on what the investigator did not understand earlier.

The training session and the first block of the first main session use *tasks* which were selected to introduce the subject to problem solving, complex mental imagery and to let them make their own experiences with introspection, self-reporting etc., but which are otherwise unconnected. The remaining sessions all use iterations of one main task: The subject is asked to cross-fold a sheet of paper in mental imagery and to imagine the resulting object in detail, while the number of folds is increased with each session. At the end of the second task block, the subject has to draw the folded sheet 2-dimensionally from its six sides.

In the remaining chapter, we will elaborate on the details of the study sketched above. Specifically, this will include the following:

- In Section 2.5, we present the *tasks* (training tasks & main task of iterated mental paper-folding),
- In Section 2.9, we discuss the *study setup* (room, equipment, software etc.),
- In Section 2.8, we present the general *procedure* followed,
- In Section 2.6, we discuss the selection of *subjects* (including an overview of all pilot subjects), and preparation of *investigator*,
- And finally, in Section 2.10, we discuss our approach towards the *analysis* of the gathered *data*.

Going into more detail, in Appendix B, we provide exact *guidelines* the investigator should follow when conducting the study. This is provided as a means to enable replication of our study, but since our study design is very demanding towards the subject as well the investigator, following these guidelines should also be considered a pivotal contribution towards ensuring reliability of our data, and hence they are included in this thesis' Appendix in unabbreviated form.

2.5 METHOD: TASKS

In this section we describe the tasks we are using in our study. There are several tasks used for a training session (cf. Section 2.8.1), and for the main sessions (cf. Section 2.8.2) there is one larger task which is divided into several sub-tasks. The original German language versions – as they were posed to our pilot subjects – can be found in Appendix A.

2.5.1 Training tasks

Overall, there are 4 training tasks divided in two blocks: In the first block, there are 3 *small tasks*, and for the second block, subjects are given *one larger, more complex task*.

First task block

For the first block of training tasks, we use three tasks adapted (translated into German) from experiments 2 & 3 in Finke et al. (1989). They are relatively simple imagery manipulation and (re-)construal tasks. The subject is given three or four short *oral instructions* consecutively and asked whether they have followed the instruction. After the last instruction the subject is asked *whether they recognise a resultant object*, e.g.:

- “Imagine the letter ‘B’.”
- “Rotate it 90 degrees to the left.”
- “Put a triangle directly below it, having the same width and pointing down.”
- “Remove the horizontal line.”
- “Can you recognise the resulting figure?” (a heart)

These tasks were chosen since (a) they are well known in the psychology literature, (b) their instructions are very clear and relatively unambiguous, and (c) their form of giving consecutive, simple instructions provide a good referential “scaffold” for a subsequent, retrospective interview. The tasks should also allow the subject to experience how objects in imagery can be re-construed, and how this may influence their memory of what was there before they recognised a figure.

Second task block

In the second task block subjects are supposed to familiarise with fashioning their own, intermittently *written self-reports* (cf. Sections 2.8 and Appendix B). Like the first tasks, it is also provided in form of consecutive instructions to construct an object in imagery, but the construction is considerably more complex, and the final question also differs. The task is taken from Wiener (2000, p.82f) (cf. also Schneider et al., 2013).

- “Imagine a square.”
- “Partition both vertical sides into three equal parts.”
- “Connect the resulting points with horizontal lines, such that we get three resulting rectangles.”
- “Now draw in both diagonals of the square.”
- “How many triangles are present in this construction?” (18)

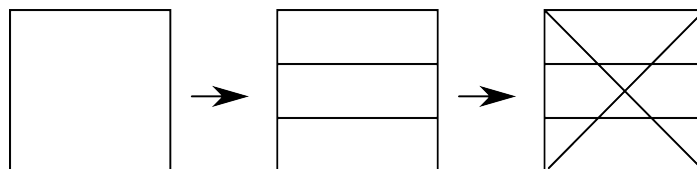


Figure 7: An illustration of the first, the second + third, and the final step of the square-task.

The task was chosen for (a) its relative difficulty (almost no subject at first arrives at the correct answer (18), which motivates them to keep imagining),

(b) its relative complexity, which allows the subject to witness the brittleness of and effort involved with mental imagery (cf. Schneider et al., 2013), and (c) once more, its step-wise instructions, which we assume ease retrospective referral.

2.5.2 Main tasks

There is one main task – *iterated mental paper-folding* – for which in every session *three sub-tasks* are presented, one in the first task block and two in the second.

First task

The first two tasks of each main session is to freely imagine cross-folding a sheet of paper a specific number of times. The first of these tasks always is a repetition of the second task of the preceding session – with the exception of the very first session, for which there is a block of “warm-up tasks” in problem solving and mental imagery (cf. Appendix A). The number of folds is increased by 1 as the sessions proceed (starting with 2 in the first session):

“Mentally fold a sheet of paper twice, such that it is halved in middle by each fold and such that the resulting creases are perpendicular to each other.

Try to acquaint yourself with imagining the resulting object.”

Beyond its main purpose, to make the subject imagine the folded sheet in detail, the exact meaning of the second instruction – “acquaint yourself with imagining the resulting object” – is deliberately kept vague. This means that the subject may continue to the next task whenever they feel they have achieved what they perceived the task to be exactly.

There are 4 main sessions, i.e. that in the last one the sheet is supposed to be folded five times. Theoretically, this could be enlarged further. The current limit of one training and four main sessions is due to the intent of fitting the study into one working week (with one session per day).

Second task

The second task is the same as the first one, but with the number of folds increased by 1. Additionally, the subject is informed, that there will be a final, third task in which *their understanding will be tested* and to which they may proceed whenever they feel ready:

“Please work on this task until you think you have become good enough at it to pass a corresponding test. If you are ready, you may proceed to the test on your own accord.

However, should you notice that you cannot do the test easily but would have to continue working in imagery, please do so and keep reporting as you did before. In other words, you should only do the test if it does not require heavy thinking on your part anymore.

Mentally fold a sheet of paper three times, such that it is halved in middle by each fold and such that the resulting creases are perpendicular to each other.

Try to acquaint yourself with imagining the resulting object.”

Third task (the test)

To make sure that the subject was imagining the folded sheet to a certain amount of detail, in the test, they are asked to draw two-dimensional sketches of the folded sheet¹⁵. The detailed instructions regarding the upcoming test (in the second task, see above) and the reminder to only start drawing if they have prepared the drawing in imagery, should ensure that the subject does not use pen and paper to improvise a solution.

“Please fashion a schematic sketch of the 3-times folded sheet, seen from left, right, top, bottom, front and back (6 2D-sketches in total). Please also mark which view is which.

Please only start drawing if you have assured yourself about each stroke in imagery before.”

If at the end of a session, the subject has not managed to draw any sketches, or if their sketches include major errors and show a significant lack of understanding of the folded sheet’s structure, the investigator conducts the following session with same number of folds, instead of increasing the number by 1 (i.e. repeat the current session).

2.5.3 Overview of sub-tasks

Overall, these are the tasks presented to the subject in the main sessions:

¹⁵ Three-dimensional sketches were considered at some point in the study design, but were judged to be too demanding with regard to the subject’s sketching capabilities.

Always:	<i>Introspect carefully and report timely but retrospectively in self-chosen short documentation breaks</i>
Session 1, Task 1:	<i>Warm-up tasks</i>
Session 1, Task 2:	<i>Imagine 2-fold</i>
Session 1, Task 3:	<i>(Plan to, and then) draw sketches of 2-fold</i>
Session 2, Task 1:	<i>Imagine 2-fold</i>
Session 2, Task 2:	<i>Imagine 3-fold</i>
Session 2, Task 3:	<i>(Plan to, and then) draw sketches of 3-fold</i>
Session 3, Task 1:	<i>Imagine 3-fold</i>
Session 3, Task 2:	<i>Imagine 4-fold</i>
Session 3, Task 3:	<i>(Plan to, and then) draw sketches of 4-fold</i>
Session 4, Task 1:	<i>Imagine 4-fold</i>
Session 4, Task 2:	<i>Imagine 5-fold</i>
Session 4, Task 3:	<i>(Plan to, and then) draw sketches of 5-fold</i>

Table 1: An overview of all tasks in the main sessions.

2.6 METHOD: SUBJECTS

2.6.1 Finding and selecting subjects

Owing to the length of the study (5 times roughly 2 hours), the difficulty of the task, and the lack of substantial extrinsic motivation (i.e. reimbursing the subjects¹⁶) it should be obvious that it is harder to find willing subjects for such a study than for a more typical half-hour psychological experiment.

This means, subjects have to be found who have (a) *enough free time* to come in for about 2 hours over 5 consecutive days (and not be exhausted from other duties), and (b) who are *intrinsically motivated* to keep thinking about a task such as ours. Obviously, not too much information about the study should be disclosed to potential subjects, hence we told them we were looking for *people who enjoy solving puzzles and riddles*.

We do not systematically assess the *prior knowledge* of subjects, since (1) the task domain is sufficiently complex that it is very unlikely someone would possess knowledge making our task substantially easier for them, (2) as long as their knowledge does not allow them to solve the task immediately, such individual prior knowledge might actually lead to an interesting perspective on the task, and (3) if any specific field of expertise would be relevant for a subject's solution of our task, this should show up in the protocol and could be thematised in the interviews. Fields of expertise we would assume could potentially be relevant for our task are e.g. Origami, advanced geometry, topology and recursive programming techniques.

¹⁶ Which would bring along its own methodological problems, since this has been repeatedly shown to decrease subjects' *intrinsic* motivation (cf. Pritchard et al., 1977)

2.6.2 Pilot cases

Table 2 shows the six pilot cases that have been studied and based on which we developed the study design and conducted the cognitive task analysis presented in Chapter 3.

Note that due to the importance of getting acquainted with the task as investigators, three of these six cases were self-studies of researchers more or less directly *involved with this study* (S4 being the author of this thesis). While the protocols gained from these cases should obviously be handled with particular care, there is no aprioric reason to not use them as exploratory data in aid of improving our study design and conducting our task analysis.

All subjects were of Austrian or German nationality, with German being their native language, in their 20s or early 30s, and except for S5, who was female, all subjects were male.

<i>Subject</i>	<i>Involved with project</i>	<i>Interviewer</i>	<i>Occupation</i>	<i>Comments</i>
S1	yes	none	PhD candidate in cognitive science	First open exploration of task, documented in Schreiber (2015)
S2	no	S4	First-year master student in cognitive science	First pilot of study design
S3	yes	S1, S4	PhD candidate in cognitive science	Increased number of tasks
S4	yes	S3	Second-year master student in cognitive science	Self-study of this thesis' author
S5	no	S4	First-year bachelor student in cognitive science	conducted with design as presented here
S6	no	S4	First-year master student in theatre studies	conducted with design as presented here

Table 2: An overview of all pilot cases.

2.7 METHOD: INSTRUCTIONS

Given the complexity of our task and the amount of responsibility ceded to the subject by letting them self-report, the instructions for our study necessarily are pretty lengthy. Acknowledging this length, the training session should provide enough time for the subject to understand and familiarise themselves with these instructions.

The instructions were synthesised from relevant literature (a) on methodology (de Groot, 1965; Ericsson & Simon, 1993; Külpe, 1920; Petitmengin, 2006), and (b) on difficulties of self-reporting (Feldon, 2010; Fox et al., 2011; Schooler & Engstler-Schooler, 1990), (c) from our own experiences gathered with the pilot cases¹⁷, and (d) from plausible reasoning – such as prefer-

¹⁷ Which means the instructions evolved significantly over the course of our pilot studies. The version presented here is the most recent, as was used for the *last two cases*.

ring errors of omission over errors of commission, since we judge missing information to be less severe than *misleading* information.

Generally, these carefully worded instructions and the training session – in which the subject gains their own experiences with trying to follow the them – form an important part of the measures we have taken to *ensure a high reliability* of the data gathered in our studies.

General instructions

The general instructions are supposed to introduce the subject to our task setup, to instruct them to work in mental imagery, to introspect, and to present them with important hints about imagery, introspection and possible difficulties of self-reporting:

The tasks you're given are no performance tests, no "measures" will be gathered, i.e. the task isn't about answering as quickly as possible. Instead, we want you to report *how* you're handling the task.

Please solve every task you're given entirely in *imagery*, i.e. without external aids. Whether you want to close your eyes during this is your choice.

Additionally, we ask you to assume an "introspective attitude", i.e. to pay attention to your own thinking. In doing so it might also happen that you notice sensory, affective or bodily aspects of your thinking.

When *reporting* your observations, it's more important to us that you report the progression of your thoughts than describing their contents as precise and eloquent as possible. In case you don't want to report certain thoughts, please mark these omissions in your protocol.

For your further orientation, in the following we'll explain some common "*pitfalls*" of composing introspective reports:

Omissions:

As with many methods of data acquisition, introspection can't register everything which might turn out to be relevant in retrospect. So don't get frustrated if an observation "escapes" you. However, as long as you still can say that "something" has happened, please report accordingly.

Inadequate identifiers:

It can also happen that one hastily and confidently describes an observation with a name/an idea, even though one already suspects that it "doesn't really fit". Also here: Better mark eventual uncertainties with "somehow" or the like, than going a bridge too far.

Verbal overshadowing:

Sometimes it may happen that you observe things in imagery that aren't easily put into words. For these cases we provide you with pen and paper for fashioning sketches.

Explaining/Hypothesising:

Especially people with a strong interest in their own thought processes are prone to get into a speculative attitude where they mix their report of what happened with hypotheses about *why* it might have happened. Should you notice that you started speculating please try to concentrate on the task again. Also please note in your protocol if you think that a specific assertion might have been of speculative nature.

Protocol instructions

In the main sessions, the protocol instructions are shown right after the general ones above (in the training session before the second task block). They are supposed to introduce the subject to *writing* their own self-reports with our study software. Special attention is being paid to problems of concurrency (and subsequent possible interference) and of using the protocol and eventual sketches as external aids.

For the following task we ask you to *autonomously* fashion a written protocol of your introspections. For this purpose there is a text input field on the left side of the next screen, in which you may type “text message-like” observations.

Regarding the exact procedure everyone has to find their own style. Nevertheless we provide the following hints to help you avoiding that working on the task and fashioning the protocol get in each others way.

Try to avoid reporting and working on the task concurrently. Instead, intermittently pause working on the task and report your observations retrospectively. Try to write *as short as possible* and *as precise as necessary*.

Usually some suitable “break points” for making pauses occur naturally, such as finishing with a sub-task or getting stuck.

Also, it might happen that you make especially impressive or fleeting observations. In these cases, please pause working on the task for a moment and make a short note.

In general, you should only pause working on the task if you have the impression that this pause will not disturb your further progress or if you have to report something important that would get lost otherwise.

If a verbal description should be difficult you also have the possibility of making small sketches of your observations (please number the sketches and use those numberings in the protocol, to make it easier for us to identify when you made the sketch).

However, please take care not to use those sketches as “tools” for thinking. To avoid that, please turn over the sheet right after finishing your sketch. Should it happen anyway that looking a sketch you’ve just drawn gives you an insight, please add an according note in your protocol.

Short instructions (reminder)

These short instructions are supposed to summarise both the general and the protocol instructions. They are shown as a reminder before the second task block.

The following tasks are no test of your performance. Some of them might be quite difficult. Please take your time while thinking and don’t get nervous if you do not know what to do at first.

Please work on the tasks *only in imagery* and assume an *introspective attitude* (also pay attention to possible sensory, affective and bodily aspects).

Separate working on the task and writing of the report. To aid this separation, please do not rest your hands on the keyboard. However, do not forget to report at all.

Pause working on the task event-contingently, e.g. after finishing sub-tasks, getting stuck or when noticing remarkable or ephemeral events.

In these pauses compose written notes ("as short as possible, as precise as necessary") or sketches about your thoughts' progress.

Please do use neither your previously written notes nor your sketches as thinking aids. Make a note in the protocol should it happen anyway.

When making your notes please pay attention to common sources of errors in introspection, namely omissions, inadequate identifiers, verbal overshadowing and speculative reasoning.

2.8 METHOD: PROCEDURE

In this section, we describe the details of the procedure for both, the training and the main sessions. Each session is divided in two task blocks and a final interview. Between these three parts, there is a compulsory pause of about 5 minutes¹⁸.

Both, the training and the main sessions involve several *debriefing interviews*, whose general procedures we describe here. The exact procedures, regarding wording, which questions to ask in which order etc. are described in the study guidelines in Appendix B.

The debriefing's main purpose is to get the subject to carefully and retrospectively elaborate on their thinking. To this aim, the interviews partly follow the elicitation interview technique described by Petitmengin (2006), which was designed to re-visit and "re-enact" memories. The main elements we adopted from this technique involve providing fitting recall stimuli by quoting either the instructions or the subject directly ("The first instruction asked you to 'Please ...'. Can you tell me what happened immediately after you read this instruction?"), and to ask open, non-suggestive questions.

Given that our task is deliberately ill-defined, and that each implementation of our study design is an individual case study, not all parts of the procedure can be as rigidly formulated as for a psychological experiment. In the end, the investigator has to use their own informed judgement, when reacting to the subject's individual behaviour, and deciding whether an error on the subject's side warrants repeating a session, or when conducting the debriefing interviews. To some extent, this contingency makes our study more similar to a Piagetian clinical interview or a Socratic dialogue than a strict psychological experiment (cf. Ginsburg, 1981; Stenning & van Lambalgen, 2004).

¹⁸ In earlier pilot cases, the pauses were optional, with the effect that at least one subject showed severe fatigue (and subsequent loss of concentration, decreased quality of self-reporting etc.) towards the end of the session.

2.8.1 Training session

As many other study designs do ours includes training, with aim to make sure the subjects have fully understood all the instructions and know what to do. However, since introspection and self-reporting cede a large amount of *responsibility* to the subjects, the training procedure and its execution must be considered with particular care.

The aims of the training session are manifold, namely to familiarise the subject with:

- Solving problems in mental imagery
- Introspecting while doing so
- Paying attention to the *progression* of thought when doing so (vs. elaborate content introspections)
- Intermittently fashioning written self-reports
- Avoiding common “pitfalls” of introspection (e.g. errors of commission, omission, speculation, ...)
- Avoiding common “pitfalls” of self-reporting (e.g. concurrency, overly vague or precise language, ...)

The exact procedure for the training session, is as follows:

1.	Welcome and explanation of the session
2.	Filling out the consent form
3.	Present general instructions
4.	First task block (with elicitation interviews)
5.	Debriefing interview about first task block
6.	Pause (5 minutes)
7.	Explain second task and present protocol instructions
8.	Second task block (written self-report, read-along)
9.	Pause (5 minutes)
10.	Debriefing interview about second task
11.	Conversation about the protocol
12.	Thanking the subject and closing the session

Table 3: Overview of the procedure for the training session.

Debriefing the first task block in the training session

The first debriefings take place after the subject finished the imagery tasks by Finke et al. (1989) (each task gets their own debriefing immediately afterwards). Their aim is to make the subject feel at home with observing their thoughts and reporting on them, as well as familiarising them with the instructions pertaining to introspection. At this point (judging from experiences in the pilots), it is to be expected that the subject’s report might contain oversights, sloppy speculations etc. We assume that letting them experience these difficulties by themselves and thematising them in this interview (as opposed to informing them theoretically) will decrease the chance of those difficulties occurring later on.

Debriefing the second task block in the training session

The second debriefing interview is based on the notes the investigator fashioned during the subject's work on the task. The aim of this interview is two-fold:

First, the second training task is the first for which the subject has written a protocol, so the interview should thematise their writing experience. Additionally, the aim of this interview is to clarify *ambiguous* or *unclear passages* in the protocol. In order to do so, as the subject recounts what they observed chronologically, the investigator may “jump in” with an inquiry as the subject reaches the relevant passage. As with the earlier interviews, the investigator should *quote the subject* and then ask whether they can provide a clarification. Unclearities can not only consist of misnomers, slips or vague descriptions, they can also pertain to an unclear chronology of events – however, judging from experience with the pilot subjects, reconstructing the correct sequence of events seems to be unsuccessful far more often than correcting other ambiguities.

2.8.2 *Main sessions*

Each main session's duration should not exceed 120 minutes. The structure of every main session is identical – the only exception is the first one, where the first task consists of a few “warm-up” tasks (cf. Appendix A). Otherwise each first task is the second task of the preceding main session. This repetition is meant to (a) allow the subject to concentrate on the task domain once more and (b) to allow us to compare their thinking about the same folding process between sessions (to see what pieces of yesterday's approach they have retained, what they seem to disregard etc.).

In each session the number of folds which the subject is asked to perform is supposed to be increased by one (starting at two, ending at five). This is conditional, however, on the subject's success in the preceding session, and the investigator may decide to let them repeat the same number of folds once over (cf. Section 2.5.2).

1.	Welcome
2.	Instructions (general & protocol)
3.	First task
4.	Pause (5 minutes)
7.	Instructions (summary)
8.	Second task
9.	Third task (sketches)
10.	Pause (5 minutes)
11.	Debriefing interview
12.	Thanking the subject and closing the session

Table 4: Overview of the procedure for the main sessions.

Debriefing in the main sessions

The debriefing interview should be conducted in a similar manner than in the second block of the training session. However, since all main session tasks are from the same task domain, the investigator must pay particular attention to the way they speak about paper-folding. Generally, they should only talk about concepts and distinctions the subject has already talked about in their protocol or in the ongoing interview. Ideally, the investigator *uses the subject's own vocabulary* for these concepts.

2.9 METHOD: SETUP

2.9.1 *Room*

Since we want to minimise the “accidental” processing of environmental cues, a neutrally furnished and quiet university room should be chosen (the actual pilot studies were conducted in different rooms, but each generally fulfilled the requirements sketched here).

The subject sits at a table, in front of a computer. The investigator is sitting close (but not intimately close) to the subject, diagonally to their left or right – such that they can see each other and have a normal, conversational distance during the interviews. Both sit in front of a computer.

Shielded from the subjects immediate eyesight (e.g. behind the computer), there are a few sheets of paper and a pen, which are available if the subject wants to draw a sketch (either during reporting, or as part of a sub-task, cf. Section 2.5.2)¹⁹.

2.9.2 *Equipment*

Beyond two computers, the equipment needed is

- a dictaphone for recording the interviews,
- an ethernet cable for connecting the two computers in a local network²⁰, and
- a paper notebook and pen (or equivalent) for the investigator's notes²¹

2.9.3 *Software*

Presentation of the instructions as well as data gathering is achieved by a *browser-based software* which was especially written for this study (based on

¹⁹ Since our task is to do with paper folding special care should be taken that no additional loose sheets of paper are visible in the subjects vicinity, and that the sheets for the sketches are not creased.

²⁰ The study software is designed in such a way that possibly privacy-sensitive data will not be transferred over the internet, but only over a local network established between the two computers directly.

²¹ Having the investigator typing notes into their computer was considered too distracting for the subject in the pilots, but this may depend on the concrete hardware in question.

common open-source components). It provides subjects with a distraction-free *full-screen interface*, where they can read their instructions, click on a 'Next'-button when they finished their current task, and type their self-reports into a text input field.

The text input is implemented based on a simple chat client (IRC), i.e. subjects enter their short, intermittent reports in a text input field, and afterwards their input is displayed above the text input field, on the screen (together with a limited number of entries added earlier, moving upwards with every entry, just like in a normal chat or instant messenger-view). Every entry which is completed by pressing return is *time-stamped* and added to a *plain-text log file*.

This adds a crucial time dimension to the protocol data, and also aids the subsequent segmentation of statements in protocol analysis. The software also adds a time-stamped message to the log file, whenever the subject presses the 'Next'-button with which they confirm finishing their current task/having read the instructions displayed etc.

The specifics of the software's use and setup are described in Appendix C.

2.9.4 Gathered data

The main source of data is an *autonomously written protocol* of the subject's introspections, which they type into our software in self-chosen documentation breaks. Additionally, they have the possibility to draw *sketches* if they deem it necessary for expression. While the subject is working on the task and intermittently writing down their observations, the investigator *reads along* on their own computer screen (with knowledge and consent of the subject) and makes notes when they do not understand something in the subject's report or want to know more about a certain passage.

After the task, the *investigator's notes* are used to guide a *debriefing interview* in which the subject is asked to help elucidate ambiguous protocol passages and further elaborate on their thinking. After the study is conducted, these interviews are *transcribed*.

In total, the data base for analysing the subject's thinking consists of their protocol, the test sketches, potential additional sketches, the investigator's notes, and the transcribed interviews.

2.10 DATA ANALYSIS: THEORY

After having described how to setup our study and how it should be conducted, we have to describe how to deal with the data gathered in the study, namely how to analyse the protocols, interview transcripts etc.

The main body of recent work regarding protocol analysis is based on Ericsson & Simon (1993)'s treatment of think-aloud studies. In these studies, verbal protocols are typically *coded*, i.e. protocol statements are encoded as belonging to different categories. These categories, or codes, are usually pre-determined by a *task analysis*, and the *hypotheses* under investigation (Chi, 1997, p.282). This way, the qualitative data from the verbal reports can es-

entially be quantified (ibid.) and used to test hypotheses with samples of large groups (cf. Someren et al., 1994).

However, since in our study, we are proceeding in an exploratory fashion, we would not like to prematurely determine a closed set of codes to which protocol statements might correspond (cf. also Newell, 1966, p.16; Newell & Simon, 1972, p.166). And since we are using a task which makes it very difficult to predict the subjects' problem spaces, it would be extremely difficult to determine such codes in the first place.

Instead, it is essential to the exploratory nature of our study that we *approach our data openly* and allow for the *possibility of serendipitous discovery*. Hence, the development of categories should be *data-driven*, and always considered preliminary.

On the other hand, we set out with a *theoretic motivation* and question, and our task has been developed not with questions about paper-folding in mind, but about *representational dynamics and problem-solving*. This means that at some point in our protocol analysis we have to apply theoretic notions or candidates for new theories, if we want to contribute to the theoretic questions we set out to solve (cf. diSessa & Cobb, 2004). This precludes us from using approaches towards protocol analysis which shun theory-driven inference completely, such as *grounded theory* (Glaser & Strauss, 2012).

The existing approaches we investigated for their possible application to our protocols thus (a) avoid coding with closed sets of categories and (b) allow the application of theoretic notions which did not emerge from the protocols at hand alone. In particular, they are Newell & Simon's *trace analysis*, Schoenfeld, Smith & Arcavi's *competitive argumentation*, and Parnafes & diSessa's *microgenetic learning analysis* (Ohlsson, 1990; Parnafes & diSessa, 2013; Schoenfeld et al., 1993).

2.10.1 Trace analysis

Newell & Simon did not name their method of protocol analysis, but Ohlsson (1990) aptly termed it *trace analysis*, the prime example being Chapter 6 of Newell & Simon (1972). The aim of a trace analysis is to interpret the verbal reports of a subject problem solving as them moving through a problem space. Such an analysis involves several steps (Ohlsson, 2012b, p.107f):

- (1) Identify the subject's problem space
- (2) Translate the protocol into a path through that problem space
- (3) Explain path by hypothesising heuristics used for deciding on the path
- (4) Verify explanation by running a computer simulation

Step (1) proceeds from the protocol and *infers* from linguistic clues contained in it, what kinds of representations, operations and goals might have been relevant for the subject, in order to postulate a representational-level problem space in which the subjects thinking can be described. The main problem for our study here is that this step presumes that there is *one* problem space and that this space *does not change* significantly for the duration of the study.

Step (2) also proceeds from the protocol, but as opposed to the first step, it mainly pays attention to the *chronological progression* of the verbal reports. Its main result should be what Newell & Simon called a *Problem-behaviour graph*. With regard to our study, this second step inherits the difficulty of the first one: If we cannot identify a fixed problem space, we also cannot describe the subject as moving through it.

Yet, the attention to the chronological progression of thinking exhibited here is also very important in our study and is what distinguishes trace analysis from other approaches to protocol analysis where chronological information is lost in the coding process (cf. Ohlsson, 1990, p.254).

Step (3) once more starts from an open interpretation of the protocols, but this time we have to identify the rules (or other forms of heuristic knowledge), which the subject used to make their decisions. While we might encounter problems similar to the ones from the steps above, this step introduces an important distinction into the analysis: Whereas steps (1) and (2) were about *describing* what the subject was doing, this step tries to hypothesise *explanations* of why the subject was doing what it was doing. This is a methodologically far more critical step, since the knowledge on which we base such decisions does not always have to be conscious, and even if it is, it may not be included in our data.

Finally, step (4) fundamentally hinges on the first three steps having been executed to a level of precision that allows a computer implementation of the identified problem space and the strategic knowledge hypothesised. Thus, while there is no reason to claim that such a thing would be impossible in our case, our understanding of our paper-folding task and the protocols gathered thus far is way too premature to call for computer simulation.

2.10.2 *Competitive argumentation*

Schoenfeld et al. (1993) describe the detailed analysis of a microgenetic single-case study of a student acquiring knowledge in a mathematical domain. The method of analysis they developed for this study was termed *competitive argumentation* (originally a term used in VanLehn et al., 1982). In their approach, they want to explain the subject's "current mental state and actions" (ibid, p.62). Every tentative explanation for a specific piece of their data they come up with, is then compared to the preceding one's, and used to extrapolate future behaviour. If this explanation conflicts with another one, decisive evidence (or counter-evidence) has to be found.

If the extrapolation of future behaviour (under the assumption that the subject's understanding does not change on its own accord) does not hold true, a justification for the apparent change has to be found. In this way of coming up with momentarily coherent explanations and reasons for their change, they fundamentally develop several alternative interpretations of their data, which they can pit against each other. At the end, the aim is to arrive at a description of the evolution of the subject's knowledge and behaviour over the course of the whole study.

Competitive argumentation thus sounds very well fitted for a microgenetic case study of representational dynamics such as ours. However, the study conducted by Schoenfeld et al. (1993) had two important advantages over ours (in terms of ease of data analysis): First, their subject used an interactive

computer system, i.e. the researchers did not only have a verbal report, but a direct recording of physical interactions, which are easier to interpret than acting in mental imagery, where we always have to rely on the indirect means of verbal reporting. Second, their task domain of graphs and equations of simple algebraic functions provides clear notations of how to speak about it, whereas in our paper-folding task, the subject has to come up with their own distinctions and terms.

2.10.3 *Microgenetic learning analysis*

Parnafes & diSessa (2013) give a detailed description of their approach to analysis. They aim their approach specifically for theory development, and distinguish three phases of data analysis:

- (1) Incubation
- (2) Theory/data negotiation
- (3) Theory appropriation

In phase (1), the researchers *openly explore the data*, i.e. they explicitly do not apply ideas from existing theories, but look for emerging patterns in the data themselves. Parnafes & diSessa characterise this phase as consisting of three activities: Observation, schematisation, and systematisation (the “OSS model”, cf. diSessa, 2008): First, we *observe* individual passages or patterns which look interesting or relevant. Then, in *schematisation*, we iteratively fashion descriptions of these observations, looking for further data validating or opposing our descriptions, and thus start generating categories of occurrences. Finally, we try to find the defining properties of these schematic descriptions and delineate their relation to other descriptive categories, arriving at a *systematic* description. At the end of the incubation phase, we should have acquired great familiarity with the data, prospective ideas about patterns to find in them, and we should have identified the main explanatory issues potential interpretations would have to deal with.

In phase (2), we use the patterns and issues identified in the first phase, to find possible existing theories which might be of explanatory value with regard to our data. The theories selected for this phase should be appropriate with regard to the researcher’s theoretical commitments, the research question and the data themselves (Parnafes & diSessa, 2013, p.27). Additionally, in order to judge whether a theory could be useful to explain our data, we might try to apply the theory in question to a short passage of our data. At the end of this phase, we should have found the most promising theory, and have identified where it might encounter difficulties in explaining our data, and could possibly have to be extended or changed.

A few theories in which we are interested are selected to be comparatively applied to the data, especially with regard to the issues identified earlier. Candidate theories are then used to generate explanations of relevant segments of the data, such that subsequently we can discuss their relative explanatory merit.

In phase (3), this theory is finally applied to our data and modified where necessary. Thus, this final phase combines theory application and theory development (ibid., p.19). Theory application should be possible at least to some extent, since the theory was identified as fitting earlier, while the

issues identified and insights gained in the incubation phase might guide theory development.

Finally, we should note that the phases and models presented by Parnafes & diSessa are of heuristic value, and generally the procedure of analysis does not follow such discrete steps. But the meta-theoretic framework they present is very well suited to our purposes, and thus we generally tried to follow their ideas in analysing our pilot data.

2.11 DATA ANALYSIS: EXPERIENCES FROM THE PILOTS

In the following we describe how the data from our pilot cases was analysed. Even though the instructions and the training session aim to prepare the subject to produce the kind of reports that would be maximally informative for us, there was a *very large variety* in style and content of the reports. This meant that each subject's report had to be dealt with individually not only on the level of its contents, but also regarding our approach towards their interpretation. This variety included differences in frequency, total number and descriptive detail of statements, in the degree of reflection/meta-cognition involved (some reports read more like Think-aloud protocols, while some were more narrative and pondering). Finally, there were also differences with regard to how much attention was paid relatively to (a) the paper-folding task, and (b) the task of introspecting and reporting²². In Chapter 4 we will discuss possible measures which might improve data quality in future studies.

Given that from our pilot runs, we could not yet synthesise a completely standardised procedure for data analysis, in the following we will only present several commonalities and peculiarities we experienced in the analysis of our pilot subjects' data.

2.11.1 Segmentation

Since our reports were already in the form of time-stamped plain text, there was relatively little work to do regarding the segmentation of statements. Usually, in the first few detailed readings of a report, we identified statements which seemed either overly short and directly connected to the subsequent statement, and those which consisted of more than one semantically not directly related sentences. To judge the "connectedness" of two statements/sentences, we considered their content as well as their time distance, assuming that the longer apart two statements were made, the less likely they should be regarded as actually being one joint statement. In the majority of cases, a sound judgement could be reached, whereas for a few doubtful cases, we opted to leave them as they were.

²² We must not forget that while we deliberately ask subjects to introspect – because of (a) its effects on learning and (b) the possibility of it producing more complete reports than Think-aloud, if dealt with carefully – this is a *task of its own* for the subjects, who therefore have to handle *two tasks*, each of which requiring cognitive capacity and bringing along the possibility of interfering with progress in the other (cf. Ericsson & Simon, 1993).

2.11.2 *Incorporation of further data sources*

As a second step, the other data sources were integrated into one common data file. These were (a) scans of any sketches drawn by the subject, (b) events described in the investigator's notes which were not reflected elsewhere (mostly about overt behaviour of the subject, or circumstantial events like distracting noise etc.), and most crucially (c) additional information from the debriefing interview.

Integration of (a) and (b) was very straight-forward as they were both generated concurrent to the report and thus can be amended to the report at the appropriate time code, but for (c) we first had to identify exactly which occasion in the report had been discussed in the interview, and then annotate this information accordingly.

This also means that the interview transcript has two separate roles, which should not be confused: On the one hand, it is an additional source of evidence about the subject's thinking which was generated *after* the written report, and hence might exhibit *change*. In this sense, it is to be analysed on its own and should be considered part of the total chronology of the subject's problem solving and learning. On the other hand, the interview discusses potential misunderstandings *about the report*, and thus contains information which is pursuant to *earlier points in time*. This information was added as a separate data column to the data file, reflecting that it does concern these statements, but should not be taken to have been generated at the same time.

2.11.3 *Annotation*

As a next step, we generally followed Parnafes & diSessa (2013)'s incubation step, i.e. we approached the amended reports openly and tried to understand what happens at each point in time. These first interpretations and issues were annotated as *commentary*, running beside each data entry. Trying to gain a closer and closer understanding of the events, this procedure was iterated many times over.

These annotations included for instance:

- The fold-number which is currently being constructed
- Paraphrases of the content of statements/events
- Open interpretative questions (to us)
- Open questions (to the subject) which might "linger"
- Identification of errors (from our perspective)
- Identification of perceived errors (from the subject's perspective)
- Explications of implicit referents (e.g. "the side" → the left side)
- Whether an idea occurred for the first time
- Whether a sub-tasks was started over again (common in complex imagery tasks, cf. Schneider et al. 2013; Schreiber 2015)

- Identification of more or less coherent periods of work on the same sub-task²³

Note that most of these annotations concerned the interpretation and reconstruction of the *chronology* of events, and not yet much distinctions about forms of representation, goals or other categories from problem-solving research or cognitive science, generally. This is due to (a) the explicit effort of suspending the application of theory, as recommended by Parnafes & diSessa, and (b) a necessity dictated by the fact that without understanding what happened when, we would hardly be able to understand by what means.

2.11.4 *Cooperative interpretation*

Where possible our data analysis was not conducted by just one person, but cooperatively²⁴. Modelled on Schoenfeld et al. (1993)'s technique of *competitive argumentation*, separate annotations (questions, paraphrases, first explanations; see list above) of one session's data were generated individually, and in a second step, these annotations were exchanged and discussed. These discussions usually resulted in agreement over a common interpretation – where no agreement could be attained, we noted the disagreement and discussed the kind of evidence that would have to be found to decide the argument (e.g. data from preceding or subsequent sessions).

The later stages of this cooperative work lead to what might be described as the schematisation part of Parnafes & diSessa's incubation phase, i.e. we started to identify and characterise *kinds* of observations and issues. It should be noted that aiding this activity is possibly the main advantage of working cooperatively: The need to communicate one's interpretation of a data entry has far larger demands on being precise in one's descriptions than if one were the only person dealing with the data.

2.11.5 *Dealing with the subjects' private epistemologies*

One problem in analysing reports of other people's introspection which we did not fully anticipate, is what we started to call their *private epistemology*. In order to fashion their reports, the subjects have to employ concepts and expressions for the psychological phenomena they are observing. These concepts and words might stem from *folk psychology*, or they could also be terms from cognitive science, academic psychology, or any other field of human endeavour in which there are expressions for psychological phenomena.

The problem is that these descriptive frameworks not only provide different linguistic labels, but they also carry with them certain beliefs, (possibly tacit) assumptions and attitudes towards one's own mental life. For instance, a "naïve" subject and a cognitive psychologist who was educated about the "imagery debate" (cf. Kosslyn, 1996; Pylyshyn, 2003; Tye, 1991) might not only mean different things when they use the phrase "mental image", they might also have different expectations about what they can do with mental images and how the mental image will "react".

²³ Largely, but not fully congruent with the actual sub-tasks posed to the subject.

²⁴ Due to a lack of time, funding and other resources, this was only possible for a small amount of our data, and mainly involved cooperation with S3 on the analysis of S6's data.

We called this set of assumptions (we assume the subject to have) the subject's private epistemology and not just their folk psychology, because we found that these assumptions and the subject's subsequent expectations do not only hinge on their ascription of beliefs, desires etc. to themselves (folk psychology), but also on their assumptions about perception, memory and the nature of knowledge (cf. also diSessa, 1993).

S6, for instance, often had the problem that he could not retrieve all the information from his memory of a mental image which he seemed to assume *should* be retrievable from this memory. A typical pattern was that he would proceed to the drawing task, expecting to be able to freely recall a mental image he had constructed earlier, and then he was surprised when this recall failed. From analysing further occurrences like this, and from his choices of wording in the reports and interviews, we inferred that he appeared to assume that (a) mental images can be inspected at will, (b) mental images can be memorised by inspecting them repeatedly, and (c) when this memory is retrieved, the image can – once more – be inspected at will. Contrast this with assumptions like “mental images must be constructed” and “to memorise a mental image, I have to memorise the procedure which I used to construct it”. It should be apparent that such differences in expectations have a large influence not only on the phrasing of the self-report, but on the progress of the task.

While this phenomenon would certainly warrant research of its own, we could only shortly touch upon it in this project. What should be clear from this discussion, is that in our data analysis we not only have to deal with a subject's individual vocabulary, but also with their tacit assumptions about the nature of the phenomena they are reporting about.

2.11.6 *Emergent annotative codes*

Continuing with the schematisation and to some extent the systematisation step, from our first interpretations of the data, we developed categories. As we already discussed earlier, we did not want to apply pre-determined codes in our analyses, so it is important to note that the codes reported here were developed from the data themselves, and that they were used as annotations and not as encodings of the complete original statements.

It should also be obvious from our discussions about case studies, the broad variety of our reports and the subject's private epistemologies that such codes have to be developed for *individual subjects*, and that we cannot assume that they turn out to be useful for other subjects' data as well. And even within one subject, notions and ideas can develop, so when defining codes, we have to be careful not to abstract away such differences.

Generally, the main value of codes were not so much to yield a full description of the data, but to help us scrutinising our interpretations. As examples for the kinds of codes that emerged from our analyses, we will describe two sets of codes in the following sections.

2.11.7 Example 1: Codes for coarse cognitive activity

The first set of codes shown in Table 5 here was developed during the analysis of S4's data (the self-study of this thesis' author). After having arrived at a first interpretation of the data, an attempt to *schematise* these observations lead to development of several categories of cognitive activity. These were mostly situated at a *coarse level of description*, i.e. they were not strictly applied to each single statement, but to mostly to groups of statements. The codes were defined to categorise abstract and task-unspecific cognitive activity. A short paraphrase of the task-related content of such a coded group of statements was annotated in addition to the code (see example in Table 6). Phenomenological descriptions about the subjective appearance of these thoughts, mental images etc. – which could also be found in the data – were largely ignored in this attempt.

While it may not necessarily make sense to pre-determine them, it is helpful to distinguish *levels of description* such as task-unspecific cognitive activity, task-specific activity, meta-cognitive activity, or phenomenological descriptions. Usually, data entries include more than one of these types of descriptions at once, so when focussing on one specific kind of description in a certain interpretation attempt, we should be aware what other kinds of information we may be ignoring.

Code	Description
@	Implicit action, doing something a specific way without active decision
M	Memory, a concrete reminding
→	First occurrence of a specific procedure/strategy
←	An implicit action has become canonised (is now the standard approach)
: –	Noticing something, Discovery of a piece of knowledge
ct	Continuing an earlier thought/idea/approach
Mct	Continuing something from memory (mem. from beyond the current session)
?!	Confusion
!!	Consciously, decisively quit doing/trying something
I?	Intrusive memory/occurrence
?–	Emerging side issue, new sub-goal
⇒	Inference, following from immediately preceding thought
::	Setting oneself a goal, intending to make a plan
()	task-unrelated side-remark

Table 5: A set of coarse, non task-specific codes developed from S4's data.

Table 6 shows an excerpt from a summary of S4's data from Session 2, Task 1 (Imagining the 2-fold). This summary describes an episode of roughly 50 minutes, pertaining to 34 protocol statements, 3 sketches and 6 relevant interview passages – hence it should be obvious that the level of coding is indeed very coarse, and mainly served the purpose of *providing an overview*.

As an example for the abstract perspective provided by this summary, consider *Entry no. 12* in Table 6 compared to the protocol segment it describes:

S4§2.16	I'm beginning with the 1-folded sheet in the established position. a "rushed" attempt at making the second fold brings forth the physical intuition that the fold won't be proper if I fold the sheet from its ends, and without paying proper attention to the crease (this idea is connected with the subsequent speculation, that this happened because I did not put the second crease into the sheet "properly", as I had done before; the feeling one would have to use a finger or even a fingernail to reinforce a crease)	ich beginne mit dem einfach gefalteten blatt in bekannter position. ein gefühlt "gehetzter" versuch die zweite faltung zu machen bringt die physikalische intuition, dass der falz nicht sauber wird, wenn ich das blatt von den enden her zusammenfalte ohne den falz sauber zu beachten (diese vorstellung ist verbunden mit der anschließenden spekulation, dass das kam, weil ich nicht wie zuvor den zweiten falz "sauber" in das blatt hineingemacht habe; das gefühl man müsse mit einem finger oder gar fingernagel einen knick kräftig verstärken)
---------	---	--

No.	Code	Summary of content
1	<i>Mct</i>	Book image from S1
2	<i>Mct</i>	Sides, areas, pairwise-ness from S1
3	: –	Corners (minor figuration of 2 sides [Vs + parallels])
4	: –	Angles (minor figuration of 2 sides [parallel + V halves])
5	: –	Lines (minor figuration of 2 sides [parallels + 1-crease])
6	: –	Marking 2-crease
7	: –	Tracing 1-crease in 2-fold (becomes V)
8	: –	Pair: bottom and top
9	⇒	Bottom is closed, hence top is open
10	?!	Confusion about Double-Vs / Nested-Vs, top/left (cf. S1-1.1.2)
11	!!	Restart construction of 2-fold
12	<i>I?</i>	Distracted by intrusion of physical intuitions
13	<i>Mct</i>	2-fold, "crow's foot" (CF) from S1
14	: –	Extended CF (figuration of 3 sides (CF + Double-Vs)
15	<i>M</i>	"Search tree" (ST), knowingly erroneous figuration
16	→	comparing Extended CF and ST
17	?–	Understand Nested-Vs in more detail, locate The ST problem
18	?–	Understand connection of Double-Vs and Nested-Vs
19	: –	Explication through tracing edges with mental finger
20	: –	Identified facing open sides of DVs and NVs as ST-problem
21	: –	"Interface" (broad figuration of 2 sides [DV + NVs])
22	?–	Further elaboration of the NVs
23	<i>M</i>	Image: someone holding the sheet with their fingers in the NVs (openness)
24	<i>Mct</i>	Tracing "lines" from 1- to 2-fold
25	: –	"Corners" become "Interface" (right corner becomes outer NV)
26	⇒	Hence, left corner becomes inner NV

Table 6: Segment from a coarse summary of S4, Session 2, Task 1.

2.11.8 Example 2: Report-type codes

Our second example is a set of codes that was developed during the analysis of S6's data. These codes were defined with a completely different aim than the ones above, and most crucially, they were not solely developed

from an examination of the data, but were in part *derived from theory*. The problem they are addressing is not so much the categorisation and description of task-level problem-solving activity, but the distinction of *different types of reporting*, which might have to be treated differently with respect to their validity (e.g. speculative reasoning), and the interpretation of their chronological occurrence (e.g. immediate vs. long-term retrospection). This also involved an investigation of the possibility to distinguish the relative amounts of meta-cognition vs. task-level cognition and the possible correlation of meta-cognition with learning events²⁵.

Ultimately, a larger scale application of these codes was not implemented, since many data entries contained mixed or ambiguous types of reporting, making it very difficult to apply the codes. Hence, in a quick trial of letting two investigators code the same report independently, an unsatisfactory inter-coder reliability of 50% was reached. However, the application of the codes was still successful with regard to one of its aims, namely to help *clarifying the chronology* of S6's thinking.

The codes were defined as follows:

Immediate Retrospection (IR):

An immediate retrospection is every introspective statement that contains a report about the immediate past. "Introspective" here is meant in a very broad sense that is not necessarily pertaining to meta-cognition (cf. MC), but to the subject elaborating on what they just have been doing and thinking. Non-introspective statements – in contrast – are unreflected statements and/or objectified reasoning (cf. CC and KE). "Immediate" spans the time from the present back to the preceding data entry.

Remote Retrospection (RR):

A remote retrospection is every introspective statement that contains a report about the past beyond the preceding data entry²⁶.

Meta-cognitive/Self-evaluative statement (MC):

Meta-cognitive statements are statements in which the subject talks about their own problem-solving, their strategies, progress etc., as well as talking about their cognitive processing in general and evaluating/appraising these statements. These evaluative statements might also include descriptions of affective states.

Speculative Self-Explanation (SSE):

Speculative Self-Explanations are statements in which the subject tries to explain their own introspective observations by constructing a potential causal model. Usually, these are triggered by IRs or MCs, in any case they relate to some observation/satement that the subject wants explained: If this is identifiable, it should be marked in an additional comment, and if it is directly contained in the current data entry, highlighted.

Intent/Plan (P):

Plans are statements about actions to be executed in the future (regardless whether pertaining to the immediate or remote future and regardless of whether this plan is

²⁵ Cf. also the coding scheme described in Berardi-Coletta et al. (1995, p.207), which introduces similar distinctions.

²⁶ It is possible that other statements like MC or SSE also are IRs, since this definition is very broad. In these cases the more specific codes (MC, SSE) should be preferred. The same goes for RR. Hence, usually IR and RR code "task-level" observations rather than meta-cognitive statements in a narrow sense.

actually executed or not). Often it is very difficult to tell whether something is a plan or a description of something which has already started to happen. It should be tried to resolve this by looking for circumstantial evidence (time difference to preceding data point, content of the following data point etc.).

Concurrent Comments (CC):

Concurrent comments are statements as are typically seen in Think-aloud protocols ("Level 1 and 2 verbalisations"²⁷): Verbalised fragments of a concurrent thought process that are unexplained and unelaborated (neither does the subject explain to themselves nor to the investigator), it is not reflected upon or put into any further context with respect to the report.

Knowledge Elaboration (KE):

Sometimes, subjects elaborate on their task-level knowledge, as if they were explaining facts and rules so someone else. This is distinct from CC since it is phrased explicitly and with a certain amount of reflection. It is also distinct from the other codes, since it is solely focussed on an "objectified" task level, without elements of introspection.

2.11.9 Summary

This concludes the description of our approach towards data analysis. After identifying the necessity to adopt an approach which does not rely on closed sets of pre-determined codes, and which allows the application of theory-driven concepts, we examined three established approaches: Trace analysis, competitive argumentation, and microgenetic learning analysis.

Drawing mostly from microgenetic learning analysis, we described our approach by reporting experiences from our pilot studies. These included the incorporation of all data sources into a common data file, the segmentation of statements, and approaching an interpretation by annotating the data co-operatively. We then discussed the issues of private epistemology, before closing the discussion with a report of two examples for emergent coding schemes that were used in the analysis of S4's and S6's data respectively.

2.12 CONCLUSION

This closes our discussion of setup and method of our study design. While the procedure evolved considerably with experiences gained in each pilot case, the state presented here is as the study was conducted for the last two subjects, S5 and S6. There certainly still are areas in which the design and its execution could be improved, and the final chapter of this thesis will contain a short discussion of possible changes that could be implemented in future work.

²⁷ Ericsson & Simon (1993) distinguish three levels of verbalisation: Level 1 are verbalisations of already present language-like thoughts, Level 2 are verbalisations of already present non-linguistic thoughts, which must be translated into language first, and Level 3 are verbalisations of thoughts which subjects became aware of through an additional process of attentive observation triggered by the task instructions.

*„Der menschliche Geist faltet als verständige Welt aus,
was implizit in seiner Wesensnatur eingefaltet ist.
In dieser Ausfaltung mißt und erkennt sich die Vernunft selbst[.]“*

— Scherbel 1999, p.196

As announced in the brief discussion of our task in Chapter 2, in this chapter we provide a detailed analysis of our task. According to Ericsson & Simon (1998, p.181), the aim of a task analysis is “(...) to explicate all of the different thought sequences and methods that participants, given their knowledge of facts and methods, could have used to generate solutions for that task.”

So, ideally, a task analysis is not merely a computational-level description of a task, but accounts for the “thought sequences” and “knowledge of facts and methods” of potential subjects. In other words, a task analysis is concerned with the identification and “charting” of the *representational-level problem space(s)* that could be used for the task in question. Yet, since for WDPs representational- and computational-level descriptions are almost identical (cf. our discussion in Section 1.4.1 of the introduction), this distinction rarely plays a large role, and many task analyses in the literature discuss their task on a purely computational level (e.g. Atwood et al., 1980, p.182f).

The reason why task analyses are conducted, is that this full knowledge of all possible solution paths and their potential influences on performance enables researchers to make predictions about their subjects: If the subject was to follow solution strategy A, their protocol should show the following characteristic types of verbalisations, their solution time should be worse than subjects using strategy B, etc.

From our earlier discussions of representational dynamics, ill-defined problem solving, and the open nature of our proposed task, it should be obvious that it is not possible to provide such an analysis for our task¹.

What we can do instead however, is to (a) approach a computational-level description of our task domain in the knowledge that it might be very different to actual subjects’ representations, and (b) describe *some* of the representational-level structures we encountered in our pilot case studies, in the knowledge that they do not exhaust all possibilities to represent the task.

While these discussions will not be sufficiently precise to enable us making predictions which we could e.g. correlate with quantitative performance measures, we consider them valuable for the following reasons: (1) They show how our task fulfils the task criteria put forth in Chapter 2, (2) They provide a framework which can be used for the interpretation of future protocols, and (3) they provide the background for a subsequent discussion of

¹ Even if we were to use a WDP, our use of introspective methods alone makes a strict task analysis such as Ericsson & Simon have in mind very difficult, since we cannot fully predict in what way asking our subjects to introspect will influence their thinking (Ericsson, 2003, p.16).

representational dynamics and possible deficiencies of problem-space theory.

This discussion is divided in four parts:

- In Section 3.1, we discuss related research from various fields of cognitive psychology (beyond problem solving),
- In Section 3.2, we present a theoretical analysis of the cross-folding domain and its properties
- In Section 3.3, we describe observations from our pilot studies, including a survey of possible ways to understand the task, typical learning trajectories, and different forms of representation we encountered,
- and finally, in Section 3.4, we discuss how these dynamics might be represented in terms of representational-level problem spaces.

Before we delve into this discussion, let us repeat the task's basic formulation (cf. Section 2.5.2). In the first session, the subject is asked to:

"Fold a sheet of paper twice, such that it is halved in middle by each fold and such that the resulting creases are perpendicular to each other. Try to acquaint yourself with imagining the resulting object."

When the subject feels sufficiently confident with imagining this folded sheet, they may proceed to a test on their own accord:

"Please fashion a schematic sketch of the 2-folded sheet, seen from left, right, top, bottom, front and back (6 2D-sketches in total). Please also mark which view is which. Please only start drawing if you have assured yourself about each stroke in imagery before."

Additionally the subject is generally asked to introspect their problem solving and to:

"(...) autonomously fashion a written protocol of your introspections. Regarding the exact procedure everyone has to find his or her own style. Nevertheless we provide the following hints to help you avoiding interference between working on the task and fashioning your protocol." (cf. Appendix A)

In each succeeding sessions the number of folds is increased by one (unless the subject encounters major obstacles, in which case the next session presents the same task again), ending with the fifth fold. For an overview of the specific sub-tasks posed in all sessions, see Section 2.5.3.

Over the course of this chapter, we will show that the task properties we defined in Chapter 2 hold for our task. They were:

- (TP1) Task requires *multiple changes* of representation
- (TP2) Changes should take place over a *longer time span*
- (TP3) The initial goal should be *vague*, allowing subjects to find their own specifications

- (TP4) Task *complexity* should be *high, but reducible* by finding better representations
- (TP5) Task should consist of several *sub-tasks*, divisible over *multiple sessions*²

3.1 RELATED RESEARCH

Even though our task is newly designed, and in the introduction we already noted a lack of research with respect to our theoretical questions, there are a few bodies of related research which should be mentioned. Since we already discussed the status quo of problem solving research in the introduction we will focus on research in other fields of cognitive science. These include mental imagery, paper folding and cognitive skill acquisition.

3.1.1 Mental imagery

Research on mental imagery has been one of the main concerns of cognitive psychology since its inception in the second half of the 20th century, and reached its peak of interest between the 1970s and 1990s (Kosslyn, 1996, p.2). The main questions asked were about its existence and the nature of its underlying representations (Kosslyn et al., 1979; Pylyshyn, 1973), and the characteristics of transformations of mental images such as *mental rotation* (Shepard & Metzler, 1971). A recent discussion concerned possible differences between rigid and non-rigid spatial transformations (e.g. rotation vs. folding) (Atit et al., 2013; Harris et al., 2013).

Most research in the field of mental imagery is *psychometric* in nature, i.e. properties of a task are sought to be correlated with characteristic performance data of subjects, such as response time, or scores. More recently, neurophysiological measures have also been taken into account (Kosslyn, 1996; Pearson & Kosslyn, 2013).

As already covered in the discussion of mental imagery in Chapter 2, there are bodies of research regarding the role of mental imagery in creativity (Finke, 1993; Kaufmann, 1990), and learning (Dunbar, 2000; Leahy & Sweller, 2008).

Additionally, there have been several studies on the role of mental imagery in solving the relatively ill-defined problem of understanding a segment of Einstein's 1905 paper "On the Electrodynamics of Moving Bodies" (Einstein, 1905). While this work seems highly relevant to us, Qin & Simon only published two small reports about these studies which did not include a detailed discussion of either their protocols or the model they developed (Qin & Simon, 1992a,b).

While research in visual and spatial imagery has formed the bulk of mental imagery research (Davies et al., 2011, p.144), more recently research on imagery in *other modalities* has emerged, such as motor imagery (Jeannerod, 1995) and acoustic imagery (Bailes et al., 2012).

² TP5 follows from the requirement that our study should follow a *microgenetic approach* (cf. Section 2.2.2, and is already fulfilled simply through the formulation of our sub-tasks and the study taking place over several sessions.

Even though it seems there is a lot research on mental imagery, there is not much we can say about our task with respect to this research. Since our task is a very open IDP, there are no sensible performance measures we could take, and hence connect to the bodies of psychometric research.

Davies et al. (2011, p.145f) propose that while most research in mental imagery has focussed on the depictive aspects of visual imagery (cf. Tversky, 2005, p.211), a large part of what they call “imagination” is not necessarily the generation of a mental image, but first of all the generation of a *scene description* (which in turn might or might not involve the generation of a visual mental image): Before we can generate a mental image, we are deciding *what* to imagine.

Since our tasks involves asking the subjects to imagine a complex and highly structured object, it might be that Davies et al.’s perspective better describes what they are doing: Subjects have to decide on a way to structure the folded sheet, find a way of structuring the “scene” they are asked to imagine (cf. Hinton, 1979). Hence, subjects are not so much concerned with generating a quasi-perceptual view, but a structured description, which only involves images secondarily³.

3.1.2 Paper folding

Within the context of research on mental imagery, there is also research specifically on *mental paper folding*. As mentioned in our general discussion of mental imagery research above, this research consists mainly of psychometric studies of well-defined paper folding tasks. These include being shown pictures of a textured cube and having to select the corresponding net (i.e. the cube unfolded to flat arrangement of six squares) from a set of possible candidates (Shepard & Feng, 1972; Wright et al., 2008), and being shown a folded sheet with a hole punched in it, and having to decide where the hole(s) appear in the unfolded sheet (Downing et al., 2005).

There is evidence, that while mental rotation and mental paper folding are partly rooted in a common skill, there are also important differences, such as mental rotation showing strong sex-related differences, while mental paper folding shows none (Harris et al., 2013). These differences are hypothesised as due to mental paper folding being a non-rigid transformation of an object (Atit et al., 2013), and due to most folding tasks involving multiple steps (as opposed to just one step in rotation), thus making the task susceptible to verbal-analytic solutions instead of a spatial-imagistic ones (Harris et al., 2013, p.107).

Beyond mental paper folding, there are also studies of people *actually folding paper*. These are mostly concerned with Origami, the Japanese art of paper folding. In a recent think-aloud study, Taylor & Tenbrink (2013) analysed the spatial language used by their subjects and concluded that their subjects “described the folding procedures spatially and in ways not mentioned in the instructions, as in ‘fold it in half’ or ‘fold the lower tip onto the upper one;’ they compared actions within and across steps or discussed the crease, object alignment, or object orientation.” (ibid., p.190). The introduction of new spatial terms and concepts will also play a major role in our task.

³ This is a highly speculative perspective, and we should be open to the possibility of large variation between individual subjects, as well as of it being wrong completely.

Additionally, there are also *formal analyses* and descriptive frameworks for paper-folding, which might be useful in describing and reasoning about the paper folding domain. This includes a computational model of Origami (Ida et al., 2015), and a qualitative description system for paper folding, especially aimed at modelling classical paper folding tasks, such as identifying the correct unfolded variant when a hole was punched into the multiply folded sheet (Falomir, 2016). While none of these systems can directly be used to model our task domain of iterated cross-folding and depictions of its folded side views, they provide a valuable background for developing our own formalisations.

3.1.3 Cognitive skill acquisition

Even though they are considered separate phenomena, research on problem solving and cognitive skill acquisition was always closely connected (cf. e.g. VanLehn, 1988). With cognitive skill acquisition being concerned with how people *get better* at mental tasks over time, and problem solving being concerned with how people solve mental tasks, we could describe cognitive skill acquisition as investigating learning, transfer and strategy change related with the repeated solving of problems from the same or similar domains, i.e. practice (VanLehn, 1996; Ohlsson, 2011, chap.6).

The main concern of research on cognitive skill acquisition is thus to reach an explanation of the *learning* or *practice curve*, describing how people develop from their first encounter with a task to full mastery (proceeding through a characteristic set of successive developments) (Ohlsson, 2012a, p.617). One approach to learn more about how these changes proceed are *microgenetic studies* which we already mentioned in Chapter 2, and in which performance measures are combined with intermittent verbal reports and trial-to-trial analysis (Siegler, 2006).

As with mental imagery, it is hard to fully connect our study to the body of research provided by cognitive skill acquisition-researchers, since we do not quantitatively assess the performance of our subjects. This is not done out of negligence, but because the ill-definedness and openness of our task (which is what we want to learn more about), do not allow the definition of quantitative performance measures (even if they did, with which task or subject properties should we try to correlate them?). Hence, our interest does not lie in describing a characteristic performance curve, or in testing the success of transfer, but – as a pre-requisite to such questions – first of all in understanding qualitatively what is involved in acquiring the skill to solve a task such as ours.

Still, since our task comprises the repetition of several sub-tasks, with the assumption that learning will be taking place, it could arguably be described as an investigating of cognitive skill acquisition as much as problem solving⁴. The reason why we still present our task in the context of problem solving is one of nuance: Rather than framing our study as an investigating of how our subjects acquire the skills necessary to be able to imagine and sketch a 5-folded sheet of paper, we propose that given their limited prior knowledge and mental resources, they have to *solve the problem of how to*

⁴ Note however that in the relatively small number of sessions in our study, subjects are not likely to achieve *mastery* in iterated mental cross-folding; but they certainly will be more proficient after the last session than at the outset.

represent the sheet such that they can succeed in our task. This perspective on what our task “is actually about” and related issues will be discussed in-depth below.

3.2 TOWARDS A COMPUTATIONAL-LEVEL DESCRIPTION: AN OVERVIEW OF THE CROSS-FOLDING DOMAIN

In the following section we will provide an overview of our task domain: The repeated cross-folding of sheets of (ideal) paper. Even though the knowledge described here is much more systematic, generalised and mathematical than everything we can expect our subjects to think of in the duration of a few 2 hour sessions, these computational-level descriptions nevertheless serve two distinct purposes:

First, a computational-level description of the task domain is a helpful tool for researchers as a shorthand and well-determined language when analysing protocols and communicating about them. Even if our interest lies in the representational structures the subjects construct to deal with our task, we often have to refer to ideal entities or processes these appear to represent. However, we should always take care not to attribute these ideal entities or processes to the subject’s thinking.

Second, such a computational-level description necessarily involves a formal apprehension of our task domain, i.e. there is a well-defined language to denote entities and processes. And while the subjects themselves may not develop such a formal apprehension of the task domain within the five sessions of our study, many of the ideas they *do* have and the concepts they *do* develop might be viewed as *precursors* to such a formal apprehension of the task domain. Hence, we might analyse our protocols with respect to the question of how such a formalised description could be grounded in the informal thinking it is preceded by. As such, especially towards the end, this discussion will include mathematical descriptions and abstractions which go beyond a basic computational-level description.

But again, it is crucial not to confuse these formal descriptions of the task domain with the representational structures subjects actually seem to employ, which generally are less coherent, tidy and language-like.

3.2.1 *The basics: Notation and varieties of folding*

The basic definition of a cross-fold as used in our task instruction is folding “a sheet of paper such that it is halved in middle by each fold and such that the resulting creases are perpendicular to each other”. Since the task is concerned with what this folding action will make the *resulting object* and its various 2D-projections (the test sketches) look like, this discussion focusses on describing the folded sheet and those properties of the folding action which will influence its appearance.

A sheet that has been folded n -times will be denoted as F_n .

Generally, a cross-folded sheet is uniquely determined by the *spatial orientation* of the unfolded sheet (“ F_0 ”), by *the number of times it has been folded*, and by the specifics of the *folding procedure* employed.

Assuming a sheet of a *normalised aspect ratio and size*⁵ and confining ourselves to *right-angled orientations and operations* only, there are *three initial positions* of the unfolded sheet: Oriented along the the sagittal, horizontal and vertical axes.

If not stated otherwise explicitly, in the following “folding” will always mean cross-folding.

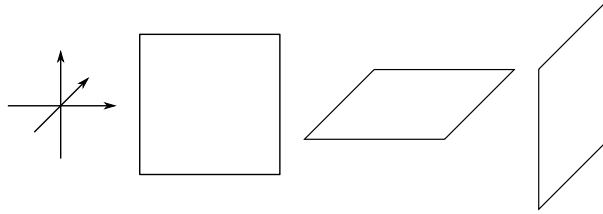


Figure 8: Three initial positions of F_0 , parallel to the three axes of Euclidian space.

There are *8 directions* in which it is possible to make a fold. The colour codes used to indicate these directions in Figure 9 will be re-used in the remaining discussion. For the first and second fold, 2 of these 8 operations yield a similarly oriented sheet (since there is not yet a different effect of folding towards or away):

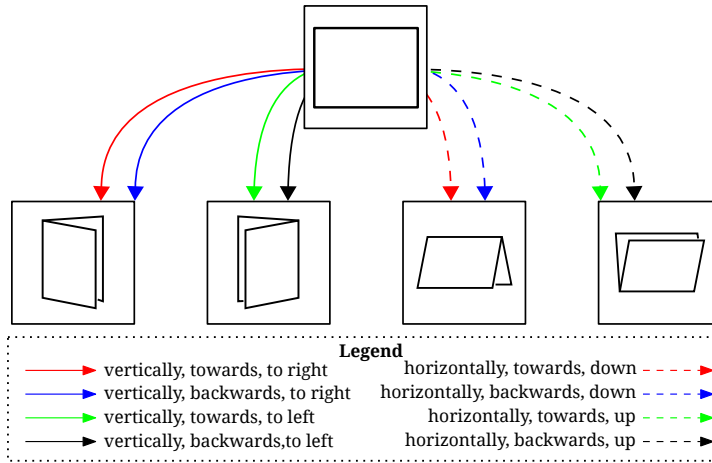


Figure 9: Eight different possibilities of folding F_0 into F_1 .

Cross-folding demands successive creases to be perpendicular to each other, i.e. after having made our first fold either horizontally or vertically, the successive fold has to be the opposite. This means that from F_1 onwards there are only *4 directions* in which it is possible to make a fold.

Alternatively, there is the possibility of *rotating* the sheet either before or after making a fold. When rotating the sheet by 90 degrees before making a new fold (or alternatively: after having made a fold), we can get a cross-fold even though we use a *fixed folding direction*.

This rotating variety has the (cognitive) advantage that the sheet remains in the same spatial orientation for each fold. Generally, there are six directions,

⁵ We do not really care about these properties and presume that they remain constant across folds. See Section 3.2.3 for a few remarks on those issues.

in which a sheet could be rotated (on each of the three axes, with an angle of either -90° or $+90^\circ$). However, only *two* of these rotation directions make sense with regard to folding, namely rotating on the plane described by the current crease and the crease about to be made with the new fold. All other rotations would lead to new spatial orientations of the sheet, making the process more complicated instead of simpler.

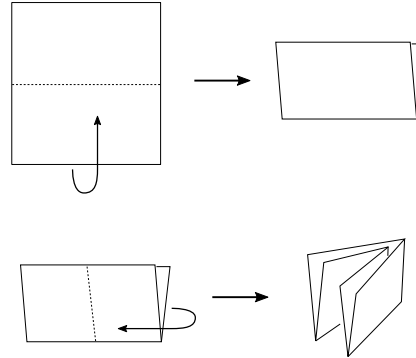


Figure 10: Illustration of a folding procedure ($F_0 \rightarrow F_1 \rightarrow F_2$) where the sheet is folded in place.

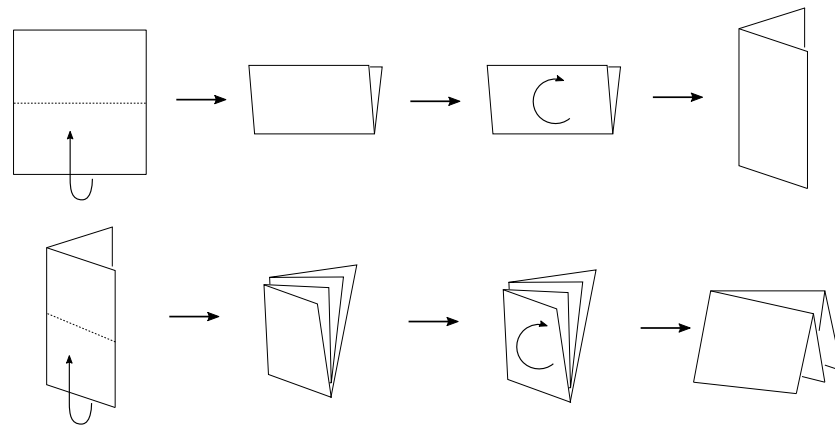


Figure 11: Illustration of a folding procedure ($F_0 \rightarrow F_1 \rightarrow F_2$) where the sheet is rotated after folding, ensuring a repeating orientation.

3.2.2 The primitive forms: The *V* and its relatives

The main change occurring between folds is not orientation, but that making a crease introduces *new vertices and edges* on the “sides” of the sheet. Hence, in order to describe a specific fold state in detail, we have to describe the configuration of edges on each of the six sides of the sheet. In order to make this easier, we will consider the 2-dimensional projections of a folded – but still slightly ajar – sheet.

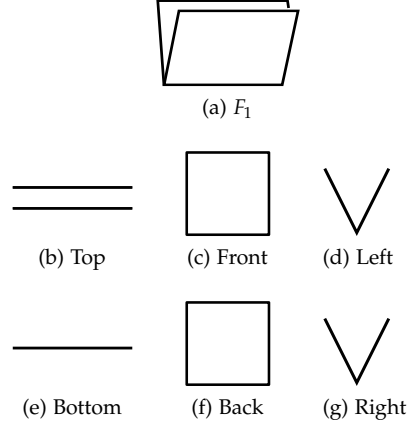


Figure 12: An illustration of F_1 together with two-dimensional projections of all of its six sides.

Even though the sides will get more complicated in higher folds, with the six sides of F_2 all the primitive forms – more complicated “versions” of which will make up all succeeding folds – have appeared (Figure 13). We name them according to their most striking respective features:

The newly made *crease* is a single stroke, the two sides parallel adjacent to the crease are *rectangles*⁶, one of the other adjacent sides to the crease is what we call a “V”⁷. The last side adjacent to the crease (opposite of the V) consists of two *nested Vs*. And finally, the last side (opposite of the crease) consists of two Vs arranged side-by-side, the *double V*. As a shorthand, we will denote the crease, V, double V and nested V of the n -th fold as I_n , V_n , DV_n and NV_n .

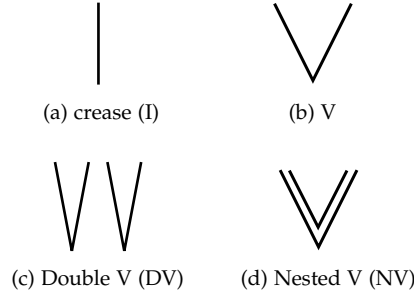


Figure 13: Schematic illustrations of four of the six sides of F_2 (leaving out the two rectangular sides). Note that their orientation does not match projections of a possible sheet configuration.

We can accurately describe a specific fold by *depicting its NV only*, together with its position on the sheet and its orientation. Using the other sides instead would not suffice to derive all remaining details needed to uniquely identify a fold.

The following figure thus precisely depicts the 8 possible variants of F_2 we can reach from the 4 F_1 's in Figure 9 (using the same colour codes for folding directions):

⁶ The two rectangular sides appear for all folds and subjects tend to ignore them rather quickly. Hence, they are not included in this and most of the following figures.

⁷ The name “V” was originally coined by C. Schreiber, cf. his report in Schreiber (2015).

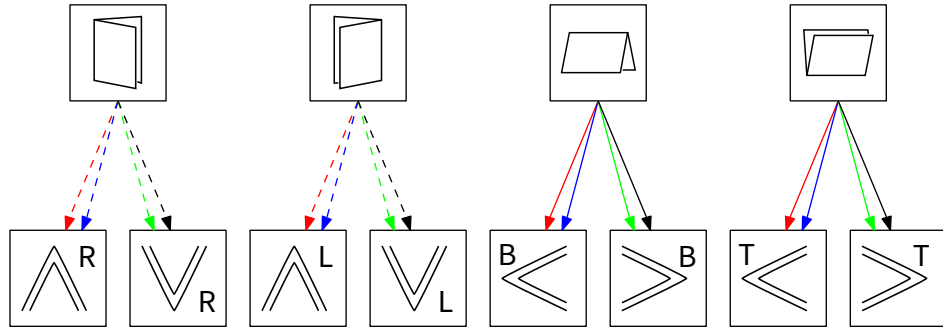


Figure 14: All 16 different possibilities of folding F_1 into F_2 . Note the different orientations of NV_2 . L, R, T, B stand for: *Left, Right, Top, Bottom* and denote the NV's position on the sheet.

3.2.3 Sheet size, fold thickness and aspect ratio

Since our task is mostly concerned with the geometry of edges on the sheet's sides, above we introduced the constraints that the sheet's size and aspect ratio will be disregarded. However, subjects might still spend considerable amounts of time working on the problems of size and aspect ratio, hence here is quick overview of the ramifications those properties introduce to the task domain.

If we use sheets with a non-square aspect ratio, each of the 3 possible spatial orientations of F_0 has a *landscape* and a *portrait* variant, doubling their number to 6.

If sheets with a $1 : \sqrt{2}$ -aspect ratio (like the ISO A series) are used, the aspect ratio will remain the same for all folds, however the orientation will alternate between landscape and portrait for non-rotating folding procedures.

When using other aspect ratios, they might change across folds. It seems unlikely, however, that a subject would do more than qualitative reasoning about which side is longer and which is shorter.

Irrespective of the aspect ratio, each fold *halves* the sheet's size along the dimension orthogonal to the crease.

Creases are introduced in the middle of the sheet, thus the length of folded edges should be halved each time we fold (although realistically, as the crease gets thicker with the increasing number of folds, more and more length of the edges will get "lost" in the "bending").

If we use a sheet of finite size, halving with each fold and the paper used has a non-zero thickness, there is an upper boundary of how many times the paper can be folded. Gallivan (2002) gives this boundary as approximately:

$$W = \pi t 2^{\frac{3(n-1)}{2}}$$

where W is the width of a square piece of paper with thickness t which has been folded n times. The maximum number for standard office paper sizes turns out to be 7, however for larger sizes up to 12 folds have been demonstrated (ibid.). For the typical thickness of office paper of roughly

0.1 mm we would require a square sheet of paper with a width of at least 29116.8 mm (29.1 metres).

3.2.4 Chirality

With an increasing number of folds, more differences between possible specific folds begin to appear. Starting with F_3 the sheet becomes *chiral*, meaning it is no longer identical with its mirror image. Instead, we get two different, mirrored versions, so called *enantiomorphs*.

Since chiral refers to “handedness”, we will call these enantiomorphs as *left- and right-handed* respectively. We define handedness with respect to the completely creased side of a fold’s NV, when oriented such that the current crease is at the bottom⁸, e.g.:

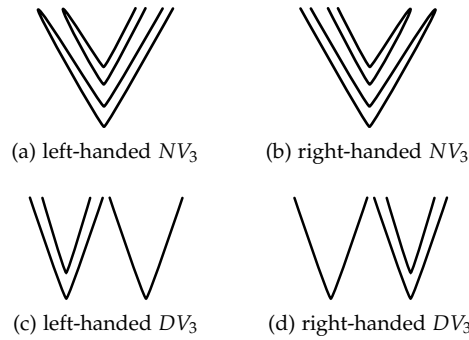


Figure 15: Two enantiomorphic versions of both complex sides of F_3 .

This chirality is caused by differences between folding away and towards⁹ and hence from each fold we can reach 4 different ones. Each fold can still be reached by 2 operations, however these do not originate from the same, but from different fold states (Figure 16).

Note that when interpreting drawings or other descriptions of the sides, differences like “mirroring” could also originate not from different folding procedures, but e.g. from idiosyncrasies in perspective-taking.

⁸ For a fold’s DV this means that it is left-handed when the nested half is on its left, and the other way around.

⁹ This dimension does not necessarily have to be away/towards, but is the one used in the exemplary space of folding directions as shown in Figure 9 and the following ones.

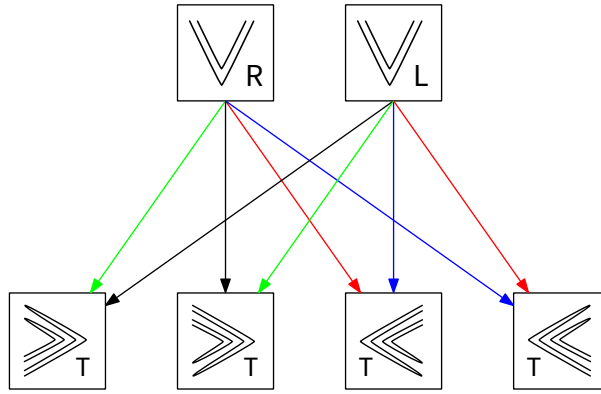


Figure 16: 8 (out of 32) possibilities of folding F_2 into F_3 . Note that each operation originating from a specific fold leads to a different new one.

3.2.5 Out-folding and in-folding

With F_4 , a more fundamental difference in the geometry of the sheet appears: Consider that there are different versions of DV_3 , and that each NV constructed by folding the preceding DV in the middle. So when constructing NV_4 , depending on which variant of DV_3 we use (cf. Figures 15d and 15c), and in which direction we make the fold, the single V of DV_3 will either become *nested within* an outer nested structure, or it will be *folded around*. Generally, we call the former version an *in-fold* and the latter an *out-fold*. We denote them with a superscript I or O, e.g. F_4^I .

So, together with chirality, which started to appear in F_3 , there are now four possible versions of NV_4 (ignoring orientation and position):

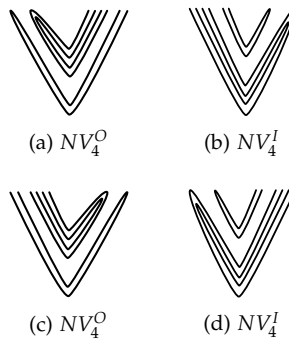


Figure 17: Out- and in-folded NVs of F_4 .

If we were to fold the sheet into different directions each fold, we could decide each time whether we wanted to get an out- or an in-fold, hence opening up the possibility of having out-folded in-folds, in-folded out-folds, with increasing numbers of possible permutations as the number of folds is increasing. So, theoretically, in F_5 there are four different NVs (disregarding orientation, position and chirality):

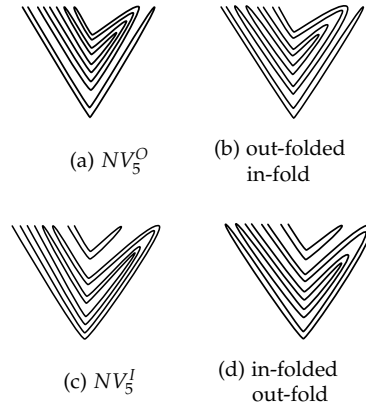


Figure 18: All possible NVs of F_5 (disregarding orientation, position and chirality).

While such “mixed” folds are technically possible, as far as we know, none of our subjects has attempted to fold the sheet like that, and given that keeping track of varying folding or rotating directions would add further demands on working memory we think it is unlikely that a subject would adopt such a procedure. Note, however, that it is still possible to produce drawings such as Figures 18(b) and 18(d) without understanding how they would come about through folding¹⁰.

So while there are more theoretically possible folds, we generally assume that a folding procedure (folding direction, rotation/no rotation, rotation direction) is chosen *once*.

For comparison, the following page shows the first five folds as they result from two different folding procedures, one with and one without rotation. In the margin, we can see a schematic 3-dimensional depiction of the two folding procedures. Beside each of these folds are 2-dimensional projections of these fold’s four sides (ignoring the two rectangular sides). The colourisation of the edges in these figures signifies *the identity of each edge*, i.e. the red edge depicting the crease of F_1 becomes a V in F_2 , then a part of F_3 ’s DV etc. It is also instructive to consider how the same edges appear at different sides of the fold in the two different folding procedures.

¹⁰ Cf. e.g. Figure 40 drawn by Subject 6.

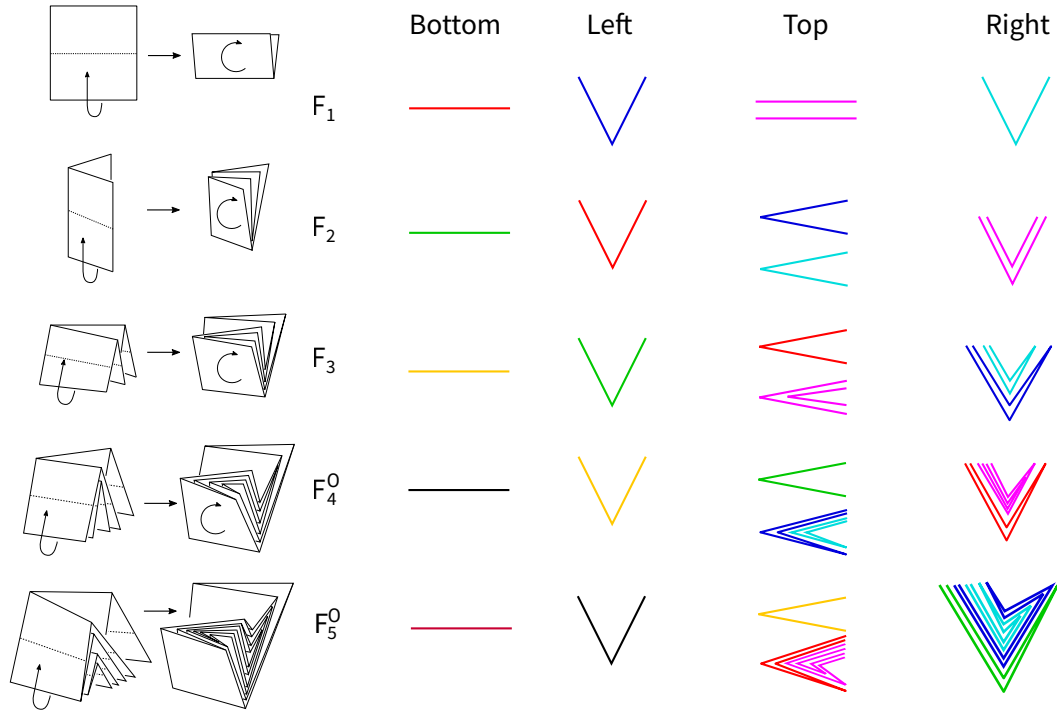


Figure 19: Configurations resulting from a horizontal-towards-upwards fold with clockwise rotation before folding (yielding out-folds), as shown in the margin. Colours trace the origin of edges over the first five folds. Note how this procedure leads to figures of the same *kind* occurring at the same position.

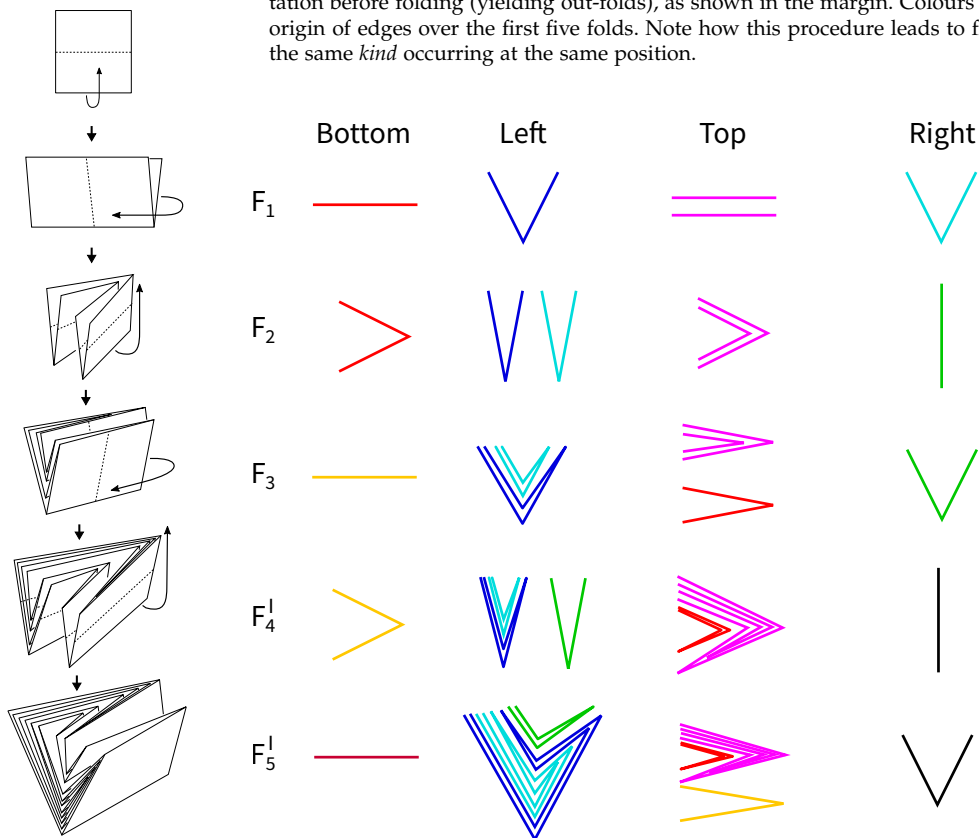


Figure 20: Configurations resulting from a non-rotating fold alternating between horizontal-towards-up and vertical-towards-to-left, as shown in the margin (yielding in-folds). Note how this procedure leads to figures of the same *origin* occurring at the same position.)

3.2.6 Folding procedures and their relation to fold states

With different starting positions, folding directions, the possibility of rotation, chirality, repeated out-/in-folding and the resulting orientation, position and detailed structure of the six sides of the sheet, we have introduced all possible sources of complexity of our task domain¹¹.

This means, there are two ways to specify a fold state: (a) by describing the details of its *folded state*¹², or (b) by describing the *process of how it was folded*.

We can describe all possible fold states formally¹³, e.g. in Backus–Naur Form (BNF), using the distinctions we have introduced up to now. We denote the NV's orientation with its "open" side (the one which opposite of the crease).

Additionally, we have to keep mind that only certain combinations of orientation and position are possible. While it does not look very elegant, the easiest way to express this in BNF is to use two different pairs of non-terminals (orientation-1/orientation-2 and position-1/position-2)¹⁴.

```

<fold-state> ::= <orientation-1> <position-1> <chirality> <in-out> <fold-num>
              | <orientation-2> <position-2> <chirality> <in-out> <fold-num>
<orientation-1> ::= "0:bottom" | "0:top"
<position-1>    ::= "P:left" | "P:right"
<orientation-2> ::= "0:left" | "0:right"
<position-2>    ::= "P:bottom" | "P:top"
<chirality>     ::= "left-handed" | "right-handed"
<in-out>        ::= "in-fold" | "out-fold"
<fold-num>      ::= "F"<positive-integer>

```

For instance the fold state determined by the following variant of the NV_5^O



Figure 21: A variant of NV_5^O .

could be described with the expression

0:left P:bottom right-handed out-fold F5.

Alternatively, however, we could describe the same fold, by specifying how it was folded. As we already know, there is more than one possibility to reach a fold state. One possibility to reach our example from above would for instance be to fold:

- ¹¹ Given the constraints introduced above, most importantly: Assuming one of three starting positions, only orthogonal states and operations and a normalised sheet size.
- ¹² E.g. signified by specifying its NV.
- ¹³ We are ignoring mixed folds for this formalisation.
- ¹⁴ This proof-of-concept formalisation makes additional simplifications: First, we skipped the trivial definition of positive integer-symbols. Second, it expresses many details about lower numbered fold states before they actually make sense (i.e. the chirality of folds $F_{\leq 2}$). Avoiding this would have meant to introduce a larger number of case differentiations, under which readability would have suffered considerably.

$F_0 \rightarrow F_1$:	vertical, towards, to right
$F_1 \rightarrow F_2$:	horizontal, away, downwards
$F_2 \rightarrow F_3$:	vertical, towards, to left
$F_3 \rightarrow F_4$:	horizontal, towards, downwards
$F_4 \rightarrow F_5$:	vertical, away, to left

If we rotated our sheet between folding, we could reach our example fold state e.g. by folding horizontal, towards, and upwards 5 times, with rotating counter-clockwise between each fold.

Leaving aside folding procedures with rotation, there are 32 *unique combinations* of folding directions to reach the example fold state shown in Figure 21, and in the general case, for any fold F_n there are 2^n possible folding procedures. This should make it obvious, that while enumerating the details of the folding procedure might be a very intuitive approach towards determining a specific fold state, there are far too many possibilities for such a description to be informative.

For a more complete overview of the relation between possible fold states and possible folding procedures, we include a state space depiction of all combinations of folding directions for procedures with and without rotation (Figures 22, 23 and 24)¹⁵. Note that Figures 22 and 23 depict different parts of the same space – for reasons of layout the depiction was simply partitioned into states where the first fold is horizontal or vertical respectively.

Note that these state spaces are included to ease understanding of the task domain and to ease protocol analysis and sketch interpretation, and not to suggest that they are problem spaces in which subjects are actually searching (unless a specific sub-problem required them to think about alternative folding procedures).

¹⁵ For the space with rotation we assume that folding and rotation directions are fixed, i.e. they are not allowed to change over subsequent folds.

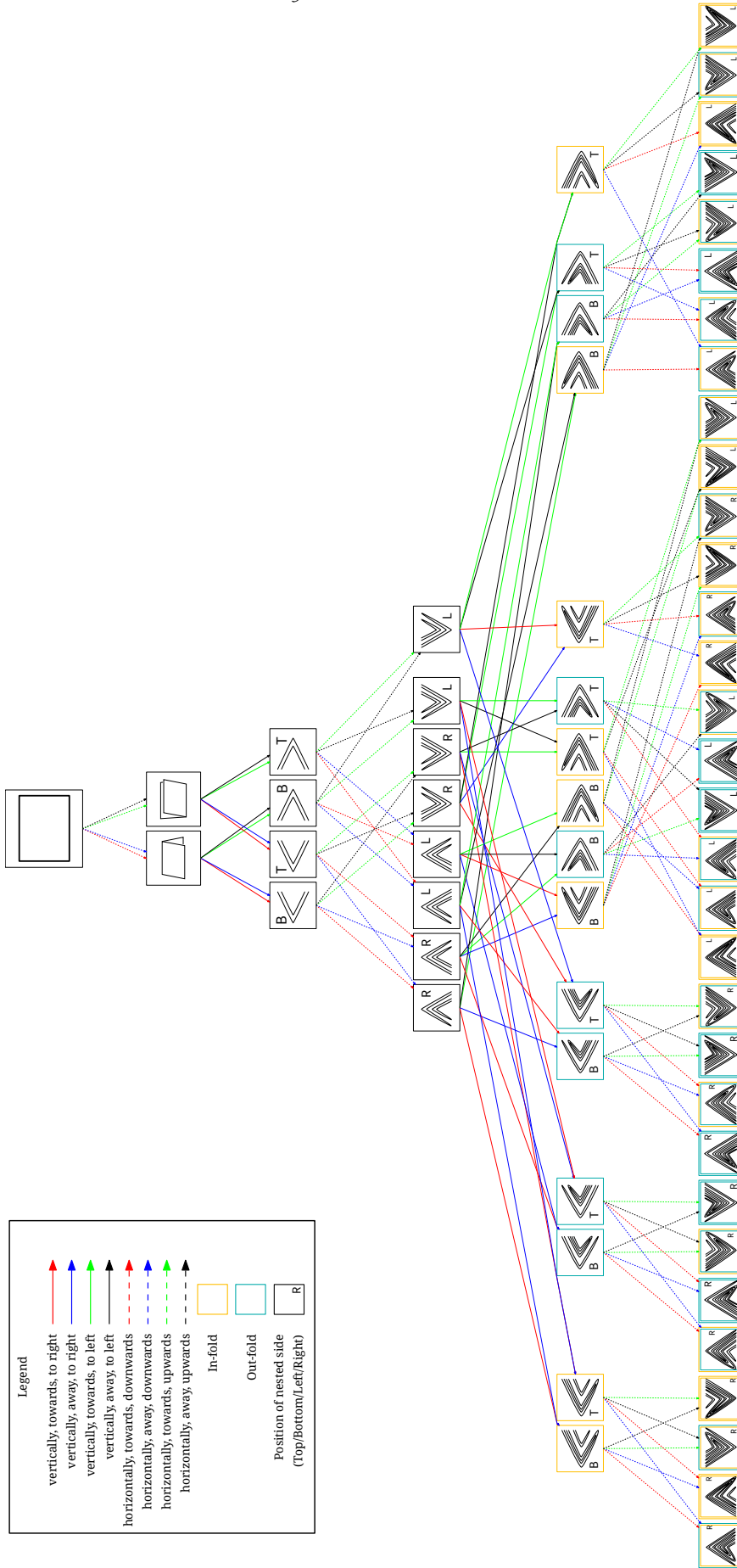
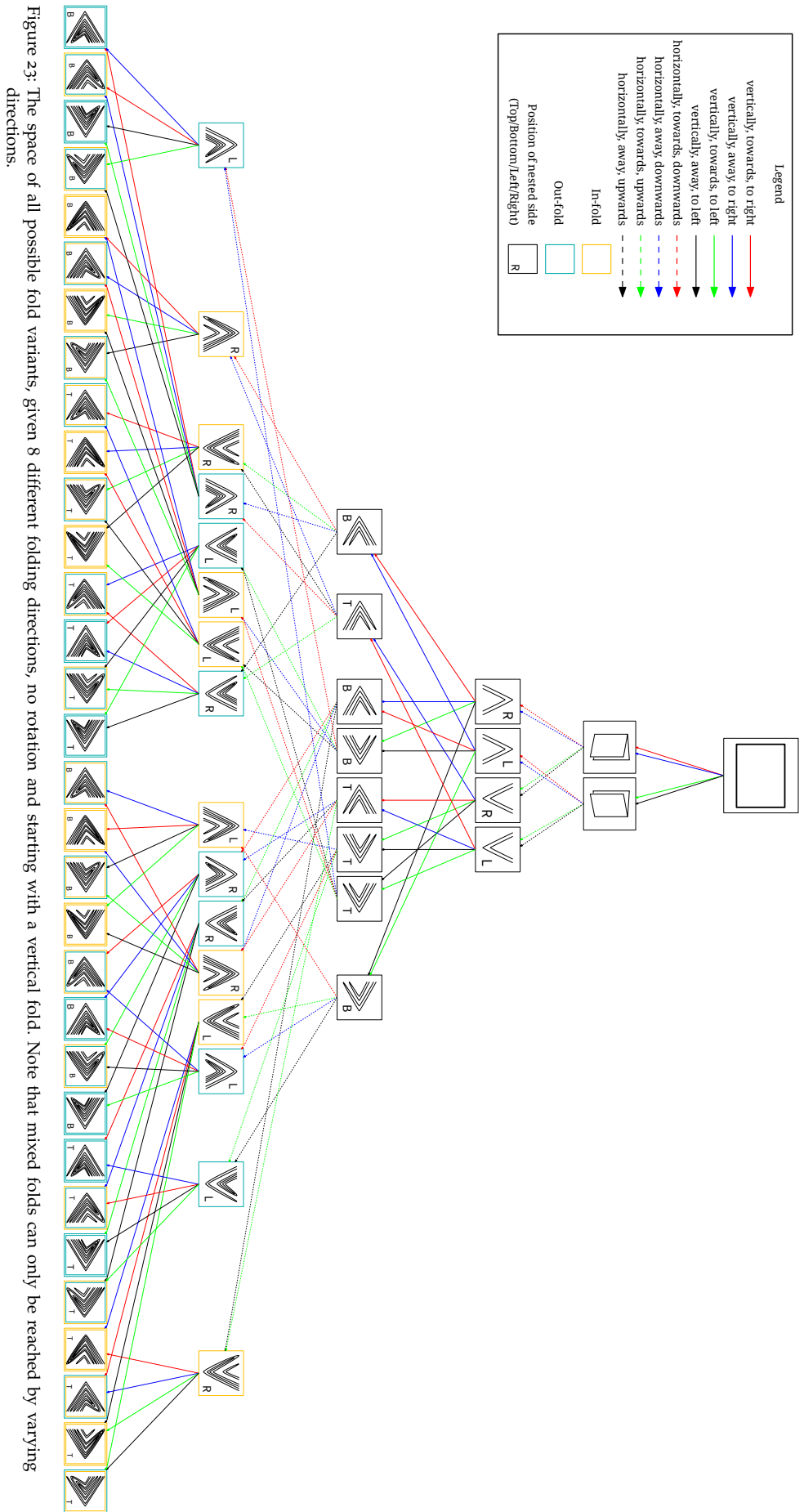


Figure 22: The space of all possible fold variants, given 8 different folding directions, no rotation and starting with a horizontal fold. Note that mixed folds can only be reached by varying directions.



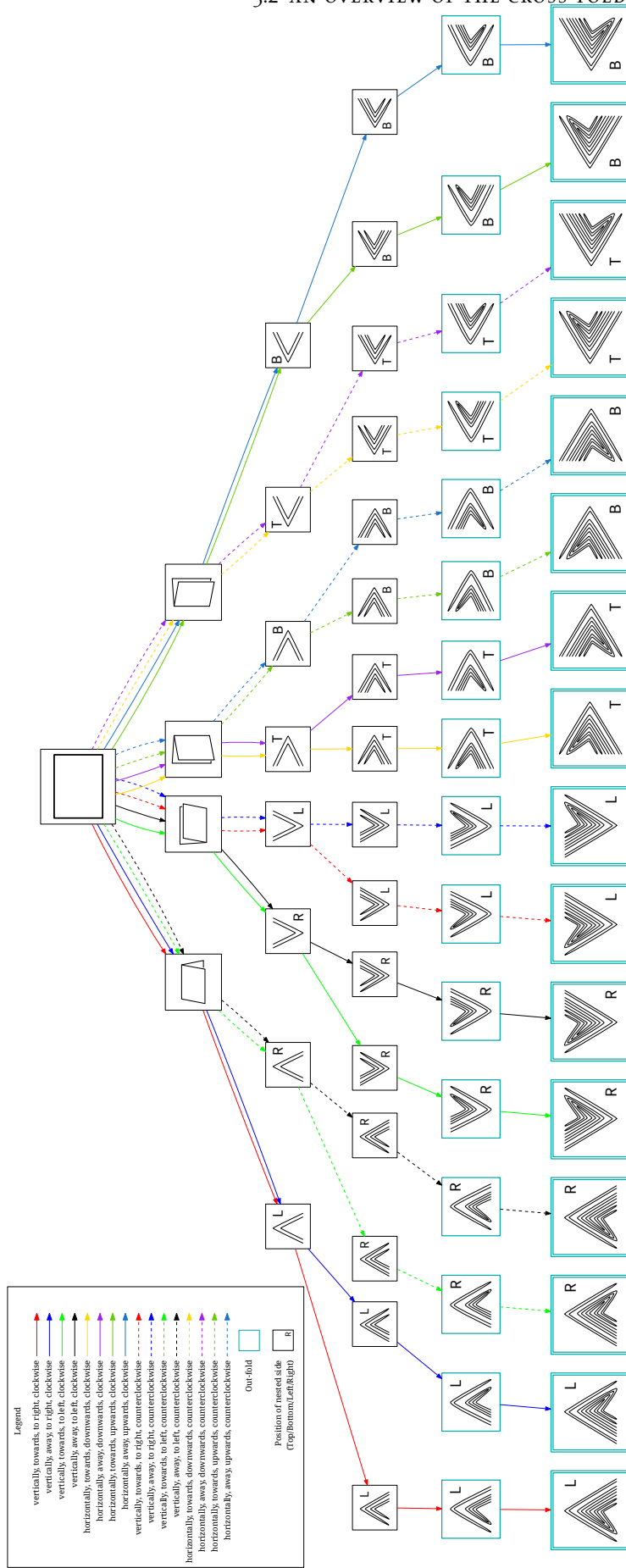


Figure 24: The space of all possible fold variants, given 8 different folding and 2 different rotation directions (totaling 16 different combined operations), with fixed operations for subsequent folds. Note that by fixing the operations, we can only reach out-folds. Allowing for *alternating* directions or rotations would yield only in-folds.

3.2.7 An overview of variants and common abstractions

We have already in passing used more or less abstract representations of cross-folds, namely ones where we ignored a side's orientation or its position on the sheet. The spaces of possible fold variants differ vastly, depending first on which folding procedures we want to allow and second which properties of the resulting folds we care about.

The *general constraints* we used for this whole chapter are that we only look at right-angled operations, we disregard the sheet's size, thickness and aspect ratio, we chose one of three starting positions for F_0 , and that if we rotate, we only rotate by $\pm 90^\circ$ in the plane of the current crease and the crease about to be made.

Optional constraints are to disallow the mixing of out- and in-folding (relevant for $F_{\geq 5}$) or to only allow out-folding (which corresponds to a folding procedure with fixed folding and rotation directions).

The *properties* – in the order in which they become relevant as the fold number is increasing – are: Orientation of the side ($F_{\geq 1}$), position of side on the sheet ($F_{\geq 2}$) and chirality ($F_{\geq 3}$).

There are 4 orientations one side can have, 4 positions on the sheet it can have (of which only 2 are ever occupied by 2 differently oriented figures, yielding 8 variants of position and orientation together) and 2 enantiomorphs.

If we disallow mixed folds, but do not abstract away other properties there is a maximum of $8 \times 2 \times 2 = 32$ variants. If we allow mixed folds, then with each succeeding fold the number of variants doubles, yielding a total of 2^{n+1} variants for the n -th fold¹⁶. If we disallow mixed folds and also abstract away orientation and position, we get $2 \times 2 = 4$ variants (e.g. the ones in Figure 17). Importantly, these 4 remaining variants are those which we cannot transform into each other by rotating and turning, but only if we were to undo and differently redo some of our folding.

Table 7 gives an overview of how many variants there are for each fold up to F_7 .

Fold	Variants				Aspect
	All	OF/IF	OF	OF/IF, no P&O	
F_1	4	4	4	1	O
F_2	8	8	8	1	P
F_3	16	16	16	2	C
F_4	32	32	16	4	OF/IF
F_5	64	32	16	4	MF
F_6	128	32	16	4	MF ²
F_7	256	32	16	4	MF ³

Table 7: The number of possible variants, given different levels of abstraction. From Left to right: All variants, only *out-* and *in-folds*, only *out-folds*, and *disregarding all aspects* of position and orientation. The right-most column indicates which variational aspect adds to the possibilities (from F_1 downwards): Orientation, Position, Chirality, Out-/In-folding, Mixed-folding, and higher-order mixed folding.

¹⁶ Making for $2^{n+1} \times 2^n$ possible folding procedures for any F_n

3.2.8 A general, recursive procedure

As was already alluded to at several points above, it is quite easy to formulate a set of *general rules* of how to produce a fold's sides from the ones of the preceding fold. Together with a fitting base case, this general step then constitutes a recursive procedure.

Since the unfolded sheet (F_0) does not contain the primitive forms¹⁷ of the folded one, we can use either F_1 or F_2 as base cases. Using only F_1 is slightly cumbersome, since it does not really have a DV yet, and since two of its sides both are single Vs, we would have to arbitrarily define one of them as its "NV" and one as the V proper. We can work around the DV problem by treating F_1 's double-stroke side like a DV (Figure 12b). In the following we will dispense with these work arounds and use F_2 as base case.

In natural language, the general step looks like this:

For all $n > 2$:

- (I) NV_n : fold DV_{n-1}
- (II) DV_n : put V_{n-1} and NV_{n-1} next to each other
- (III) V_n : fold I_{n-1}
- (IV) I_n : is always an I

By specifying the direction in which to fold for rules (I) and (III) and whether "next to each other" means "to the left of" or "to the right of", we can generate either out- or in-folded sides. Note that this procedure is only concerned with generating the individual sides and completely ignores orientation and position.

Rules (II)–(IV) are easy enough to follow, but we ought to explain how exactly rule (I) – folding a DV into an NV – is supposed to work. Let us look at an example:

Here, we fold a DV_4^O into an NV_5^L . The 4th crease is on the opposite side of DV_4^O , in between the two sub-parts (the dashed line in Figure 25). Hence the 5th crease has to be orthogonal to that, horizontally "through" DV_4^O (the solid line). The lower parts of DV_4^O are folded upwards and to right, being turned on their head in the process.

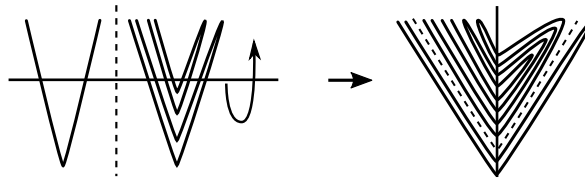


Figure 25: Out-folding a DV_4^O into an NV_5^L . The 4th crease is indicated with a dashed line, the 5th one with a solid line.

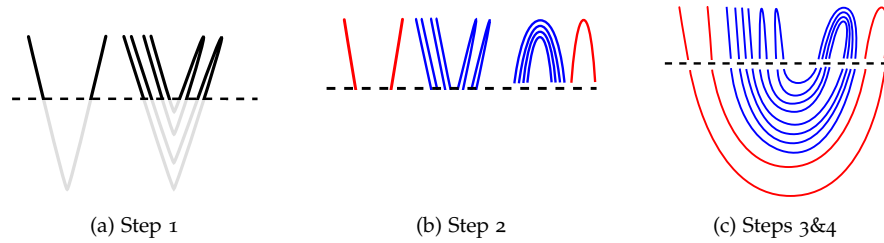
Still, it is hard to directly see and draw the nested structures (which in our example are landing on the right-hand side of the figure), so let us consider the following procedure of how to explicitly construct them (cf. Figure 26):

1. Draw the upper half of the DV¹⁸

¹⁷ The four sub-figures shown in Figure 13.

¹⁸ "upper" with respect to the orientation in which we usually depict sides here, e.g. in Figure 25. Many subjects called these sides of Vs "open" as opposed to the creased, "closed" side.

2. Rotate the lower half by 180° and put it next to the result of step 1
3. Close the bottom of this figure by joining the leftmost edge with the rightmost one in a large arc
4. Proceed drawing these arcs from the outside towards the middle, until all edges are joined up

Figure 26: Constructing NV_5^O from DV_4^O .

The rotation in step 2 can be divided further into two steps which are less simple to describe, but simpler to follow:

- To the right of the result of step 1, put half as many upside-down nested Vs as there open edges on the nested side of the DV^{19}
- To this figures' right, put a single upside-down V

3.2.9 Barbra notation

If we want the recursive procedure above to produce more than a natural-language description of a folded sheet's sides, we need a *syntactic representation* of the sides on which a formalised procedure can operate.

The representation used here is – like the NV-construction procedure above – based on the observation that we can depict an NV by enumerating the “tips” of the edges at the open side of the figure and by noting whether they are simple edges, Vs or nested Vs. As we did above, the full side can be reconstructed when we “close” these edges by joining them up with concentric arcs (joining the left-most edge with the right-most one and so on, cf. also related drawing procedures invented by subjects as described in Section 3.3.7).

In order to allow easy use of string-based algorithms, we chose an ASCII character-based notation, where open edges are represented by “|” (a horizontal bar) and Vs and their possible nestings by “()”, i.e. pairs of matching parentheses²⁰. Since the DV actually comprises two disjoint sub-figures, we use the symbol “;” to denote the gap between separate sub-figures.

For its use of *bars* and *brackets* the notation was dubbed *Barbra notation*. Table 8 gives a few examples:

¹⁹ We need half as many nested Vs as open edges, since every V has two edges on its open side.

²⁰ While using backwards and forwards slashes would have been a more iconographic choice to depict the tips of Vs, using parentheses allows to make use of the facilities for checking parenthesis matching that many computer environments provide.

Name	Barbra	2D drawing
DV_3	;	
NV_3	()()	
DV_4^O	; ()()	
NV_4^O	()()()	

Table 8: Three different representations of the same side of a folded sheet

3.2.10 A recursive algorithm

Using the recursive procedure together with Barbra notation, we can finally formulate a recursive algorithm, able to generate the n -th fold's DV and NV:

Algorithm 1 FOLD(n) recursively generates Barbra notation-based representations of the n -th out-fold's DV and NV

```

1: function FOLD( $n$ )
2:   if  $n = 1$  then
3:      $DV_n \leftarrow "||;"$ 
4:      $NV_n \leftarrow "||"$ 
5:     return  $DV_n, NV_n$ 
6:   end if
7:   if  $n = 2$  then
8:      $DV_n \leftarrow "||;||"$ 
9:      $NV_n \leftarrow "|||"$ 
10:    return  $DV_n, NV_n$ 
11:  end if
12:  if  $n > 2$  then
13:     $m \leftarrow n - 1$ 
14:     $DV_m, NV_m \leftarrow \text{FOLD}(m)$ 
15:     $DV_m^{left}, DV_m^{right} \leftarrow \text{SPLIT}(DV_m)$ 
16:     $num\_parens \leftarrow \text{LEN}(DV_m^{right})/2$ 
17:     $Vs \leftarrow \text{NESTED\_PARENS}(num\_parens)$ 
18:     $DV_n \leftarrow "||;" + NV_m$ 
19:     $NV_n \leftarrow DV_m^{left} + DV_m^{right} + Vs + "()"$ 
20:    return  $DV_n, NV_n$ 
21:  end if
22: end function

```

Lines 2–11 cover F_1 and F_2 , and lines 12–21 cover the recursive step. In natural language, the recursive step consists of the following:

- Generate DV_m and NV_m by calling fold again (with $m = (n - 1)$)
- Split DV_m at its semicolon
- Assign num_parens the number of half the length of the right side of DV_m
- Generate num_parens -many nested parentheses
- Concatenate " $||;$ " and NV_m to get DV_n
- Concatenate left and right side of DV_m and the generated parens and " $()$ " to get NV_n

To get an in-fold instead we have to account for the *left* half of the DV to be the nested one, and to put the single V around the nested structure for the NV , not the other way around. This means we have to change lines 16, 18 and 19 into:

```

num_parens ← LEN( $DV_m^{left}$ )/2
(...)
 $DV_n$  ←  $NV_m$  + " $||;$ "
 $NV_n$  ←  $DV_m^{left}$  +  $DV_m^{right}$  + " $()$ " +  $Vs$ 

```

where DV_m^{right} always equals " $||$ ", and we thus trap a singly-folded V – " $||()$ " in Barbra – *within* an outer, nested figure (as opposed to in an out-fold, where DV_m^{left} would equal " $||$ " and the singly-folded V thus wraps around the nested one).

We can also automatise the completion of a Barbra notation-string into a schematic drawing, by joining up the string with Unicode-based line characters:

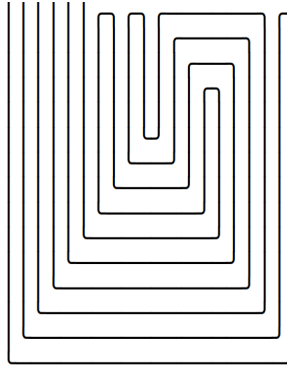


Figure 27: A Unicode-based rendering of NV_5^O

3.2.11 A simpler, iterative procedure

Using the syntax to describe different folds and the recursive algorithm to generate them, we might notice certain repeating patterns in successive folds more easily than if we were to look at hand-crafted drawings. Most strikingly there are interesting regularities in the number of nested V s.

Let us look at the first 12 in-folds in Table 9 (printed sideways on the following page).

It appears that the maximum number of nestings in successive folds increases by the doubled-up powers of two (1, 1, 2, 2, 4, 4, ...), so the number is described by the partial sum of this sequence²¹:

$$1 + 1 + 2 + 2 + \dots = \sum_{i=0}^n 2^{\lfloor \frac{i}{2} \rfloor}$$

As a shorthand we will call this partial sum *foldseq*. The first elements of *foldseq* are:

$$1, 2, 4, 6, 10, 14, 22, 30, 46, 62, \dots$$

Let us look at the pattern of successive nestings again, but introduce a simple run-length encoding to notate nested Vs to make them easier to recognise:

Side	RLE-Barbra notation
NV_1	
NV_2	
NV_3	(1)(1)
NV_4^I	(1)(2)
NV_5^I	(1)(1) (1)(4)
NV_6^I	(1)(2) (1)(6)
NV_7^I	(1)(1) (1)(4) (1)(10)
NV_8^I	(1)(2) (1)(6) (1)(14)
NV_9^I	(1)(1) (1)(4) (1)(10) (1)(22)
NV_{10}^I	(1)(2) (1)(6) (1)(14) (1)(30)
NV_{11}^I	(1)(1) (1)(4) (1)(10) (1)(22) (1)(46)
NV_{12}^I	(1)(2) (1)(6) (1)(14) (1)(30) (1)(62)

Table 10: The first 12 in-folds' NV in Barbra notation with a run-length encoding abbreviating the number of nestings.

From Table 10 we can see that *foldseq* does not directly describe the number of nestings used in each fold, but that

- each nesting is always preceded by a two bars and a single V (i.e. a "1")²², and that
- depending on the *parity of the fold number* only every other element of *foldseq* is used, starting either with the odd ones (1, 4, 10, ...) or the even ones (2, 6, 14, ...).

In the odd-numbered NV_{11}^I for instance, the successive nestings visible are

$$1 - \mathbf{1} - 1 - \mathbf{4} - 1 - \mathbf{10} - 1 - \mathbf{22} - 1 - \mathbf{46},$$

whereas for the even-numbered NV_{12}^I they are

²¹ which is sequence A027383 in the On-Line Encyclopaedia of Integer Sequences (OEIS, 2016).

²² These are the multiply folded single Vs from preceding folds' DVs, which in an in-fold always are "surrounded" by the nested Vs.

Side	Barbra notation	max. # of nestings	Diff.
NV ₁ ^I		0	-
NV ₂ ^I		0	0
NV ₃ ^I	()()	1	1
NV ₄ ^I	()()()	2	1
NV ₅ ^I	()() ()((()()))	4	2
NV ₆ ^I	()() () ()((()()))	6	2
NV ₇ ^I	()() ()((()())) ()(((((()()))))))	10	4
NV ₈ ^I	()() () ()((()())) ()(((((()()))))))	14	4
NV ₉ ^I	()() ()((()())) ()(((((()())))))) ()(((((()()))))))	22	8
NV ₁₀ ^I	()() () ()((()())) ()(((((()())))))) ()(((((()()))))))	30	8
NV ₁₁ ^I	()() ()((()())) ()(((((()())))))) ()(((((()())))))) ()(((((()()))))))	46	16
NV ₁₂ ^I	()() () ()((()())) ()(((((()())))))) ()(((((()())))))) ()(((((()()))))))	62	16

Table 9: The first 12 in-folds' NV in Barbra notation and the largest number of nested Vs in each one. The relation between the numbers might at first not appear regular, until one looks at their successive differences (rightmost column).

$$1 - 2 - 1 - 6 - 1 - 14 - 1 - 30 - 1 - 62.$$

When we look at the elements of *foldseq* together with their indices, we can see that the number of nestings in *odd-numbered* folds correspond to the *even-numbered* elements *foldseq* and vice versa.

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	...
Element	1	2	4	6	10	14	22	30	46	62	94	126	190	...

Table 11: The indices and respective elements of *foldseq* (OEIS A027383), which governs the number of nestings in the NV.

Looking at our examples NV_{11}^I and NV_{12}^I again, we can also see that the last element of the sequence they use are the 8th and 9th respectively. Generally, the last element of the sequence used for the n -fold is the $(n - 3)$ -th.

So, the NV_n^I consists of (starting with F_3):

- $||| + ||() + \text{foldseq}(1)\text{-many parens} + ||() + \text{foldseq}(3)\text{-many parens} + \dots + ||() + \text{foldseq}(n - 3)\text{-many parens}$, for even n
- $|| + ||() + \text{foldseq}(0)\text{-many parens} + ||() + \text{foldseq}(2)\text{-many parens} + \dots + ||() + \text{foldseq}(n - 3)\text{-many parens}$, for odd n

With this information, we can now formulate a *non-recursive* algorithm for generating NV_n^I :

Algorithm 2 FOLD(n) non-recursively generates Barbra representations of the NV of the n -th out-fold

```

1: function FOLD( $n$ )
2:   if  $n = 1$  then
3:     return "||"
4:   end if
5:   if  $n > 1$  then
6:      $NV_n \leftarrow "||"$ 
7:     if  $n \bmod 2 = 0$  then
8:        $NV_n \leftarrow NV_n + "||"$ 
9:       for  $i \leftarrow 1; n - 3; i + 2$  do
10:         $NV_n \leftarrow "||" + NV_n + \text{NESTED\_PARENS}(\text{FOLDSEQ}(i)) + "()"$ 
11:      end for
12:     end if
13:     if  $n \bmod 2 = 1$  then
14:       for  $i \leftarrow 0; n - 3; i + 2$  do
15:         $NV_n \leftarrow "||" + NV_n + \text{NESTED\_PARENS}(\text{FOLDSEQ}(i)) + "()"$ 
16:      end for
17:     end if
18:     return  $NV_n$ 
19:   end if
20: end function

```

To get an in-fold we have to change lines 10 and 15 into:

$$NV_n \leftarrow NV_n + "||()" + \text{NESTED_PARENS}(\text{FOLDSEQ}(i))$$

Algorithm 2 only yields the NV, but we can easily construct DV_n^O , by concatenating $||$; with $NV_{(n-1)}^O$ (or the other way around for DV_n^I).

3.2.12 A mathematical description of NV and DV

Based on the Algorithm 2, we can also describe NV_n and DV_n in mathematical rather than algorithmic notation.

$foldseq$ is known by now, $nest(n)$ is the single step in the body of the for-loop of Algorithm 2, (a_n) yields the even/odd-numbered indices depending on n 's parity and m tells us how many iterations of $nest(a_n)$ there are, also depending on n . The following is a description of the in-fold:

$$\begin{aligned} foldseq(n) &:= \sum_{i=0}^n 2^{\lfloor \frac{i}{2} \rfloor} \\ m &:= \lfloor \frac{n-3}{2} \rfloor \\ (a_n) &:= \begin{cases} (2k)_0^m, & \text{for } n \bmod 2 = 1 \\ (2k+1)_0^m, & \text{for } n \bmod 2 = 0 \end{cases} \\ nest(n) &:= "||()" + nested_parens(foldseq(n)) \\ NV_n &:= \begin{cases} "||", & \text{for } n = 1 \\ "||| + \sum_{i=0}^m nest(a_i), & \text{for } n \bmod 2 = 1 \wedge n > 1 \\ "|| + \sum_{i=0}^m nest(a_i), & \text{for } n \bmod 2 = 0 \wedge n > 1 \end{cases} \\ DV_n &:= NV(n-1) + ";||" \end{aligned}$$

3.2.13 Conclusion

This concludes the formal description of our task domain. The reader should now be equipped with enough knowledge about cross-folding to understand the following sections, which are based on subjects' protocols and our interpretations of them.

While it might have been possible to define a computational-level problem space for our task, we refrained from doing so, since the vastly different

levels of abstraction, and the varying amount of knowledge necessary to apply them in describing folds could hardly be expressed in a single problem space. So, if we were to define such a single computational-level problem space, it would very often either smooth over distinctions relevant to a subject, or it would require the specification of details about which a subject does not (yet) possess any knowledge. Therefore, instead of a single problem space, we provided a series of descriptions, abstractions and formalisations which all are relevant for different aspects of the task (domain).

This preceding section should also have demonstrated to the reader the potential mathematical complexity of our task domain (even though at first it sounds relatively simple) and the means by which this mathematical complexity can be described by identifying and formalising various domain principles²³. Thereby our task fulfils *task property (TP₄)* (high, but reducible complexity). The large number of folding procedures, variants etc. which all are possible specifications of our task instructions should also provide a first indication with respect to *task property (TP₃)* (vague goal).

²³ Note that these formal descriptions are very far away from an intuitive apprehension of the physical activity of paper folding. Whereas the recursive algorithm still embodies some abstract principles of the actual folding process, the computationally simpler, iterative algorithm relies completely on a description of the structure of the Barbra strings which is completely detached from folding.

3.3 OBSERVATIONS FROM THE PILOT STUDIES: REPRESENTATIONAL DYNAMICS AND LEARNING TRAJECTORIES

In the following we present an overview of our observations taken from our six pilot studies. Since we cannot predict all representation-level structures and strategies a subject might possibly employ, our observations should be taken to possess *heuristic* value. The language used to describe the cross-folding domain largely draws on our formal analysis above, although we take care to distinguish between the abstract formal entities and the subject's subjective notions.

The analysis is structured in sections of roughly chronological order, aggregating developments with broadly related themes. Note that through this aggregation, the descriptions given here might suggest a more teleological progression of a subject's thinking than would be warranted in a concrete case. It might e.g. also happen that even though a subject has progressed to newer, more complex ideas, they sometimes "arbitrarily" return to using their earlier ideas, that they repeat mistakes which they had already resolved, and so on. Generally, the following descriptions do not constitute stable stages of development, but are ordered heuristic tools in aid of analysing concrete cases.

Each section describes related aspects of our pilot subjects' thinking which we were regarding either as typical or remarkable. This includes a general introduction, several examples, illustrations, and quoted passages from individual subject's reports. Along with the representations and strategies employed by subjects at different stages in the task, we try to describe their current task understanding, i.e. the goals or sub-goals the subject is explicitly or implicitly (according to our interpretation) trying to achieve.

These descriptions should enable us to discuss the possibilities of defining representational-level problem spaces in which these activities could take place in Section 3.4.

3.3.1 Preliminaries: Repetition, confusion and mind-wandering

As mentioned above, for reasons of understandability the following descriptions of different ways to represent and work on our task are pretty terse. Therefore they might obscure observations which were commonly made in analysis of our pilot data, but which are hard to depict in such a presentation. Mainly, this concerns the general abundance of confusion, mind-wandering, losing one's track (especially losing a mental image) and of repeating and starting over again some mental processes.

Even given the large variety between the representations, strategies, and styles of reporting between our pilot subjects, these phenomena could be found in all subjects. Even when a protocol entry reads "I'm imagining X and can see it clearly", it might be the case that a few seconds later the same subject reports that the image and the information they hoped to gain from it has been lost again. Seen from a strict information-processing perspective repeating an exact same process does not really make sense, since it should not yield new information²⁴, so we might assume that either the process

²⁴ Interestingly, Newell & Simon still already allowed for the possibility of repeating the same operation in their formal descriptions of problem-solving (cf. Newell (1967, p.10f).

had originally failed and is executed once more to try again, or that there is actually variation in these repetitions: Be it in the sense of practise, with some sort of learning process running “on” these repeating instances, or in the sense of shifts of attention towards different aspects of the same process.

But even without a hypothesis about the functionality of repetition, we have to acknowledge its abundance: Our problem-solving protocols are full of repetition, meandering, confusion, fatigue and sometimes even flights of ideas (cf. Liepmann 1904). It is important to keep in mind the possibility that these phenomena might not be symptoms of an imperfect or malfunctioning system, but that they actually contribute functionally to the system’s aims. While it is beyond the scope of this thesis to assess this hypothesis, the re-narration of his cross-folding observations in Schreiber (2015) is paying special attention to these rarely considered aspects of thought.

3.3.2 *Preliminaries: On the use of folk-psychological terminology*

In line with Parnafes & diSessa (2013)’s approach to data analysis, we tried not to introduce too many exogenous theoretical notions into our descriptions, and instead to remain relatively close to the data. However, we still have to employ some theoretical terms when describing our observations. The terms in question are common, folk-psychological notions such as *analogy*, *metaphor*, *(mental) image*, *association*, and the distinction between *implicit* vs. *explicit*. While the exact meaning of most of these terms is contended in cognitive science, their “broad” meaning allows us to fit them naturally to the (for the most part) equally folk-psychological reports by our subjects. This means that our usage of these terms should also be taken in the broadest, folk-psychological sense, and not to embrace specific theories from within cognitive science which try elucidating these terms.

Ultimately, finding a language which appropriately describes data such as ours is still an open question. The subject’s language often includes a mixture phenomenal and functional descriptions with no clear delineation between them, and depending on their own tacit assumptions towards functionality, they might ignore important parts of their thinking or include confusing amounts of detail (see our discussion of the subject’s private epistemologies in Section 2.11.5). Ideally, studies like ours will also help to improve the terminology which is used to categorise and communicate about such reports of thinking, but for now we have to rely, at least partly on folk-psychology. There already are efforts to improve upon folk-psychological terminology (e.g. Raab & Eder, 2015), but they are not yet sufficiently well established.

3.3.3 *Preliminaries: Overview of sub-tasks*

In order to easily refer to a specific sub-task, we repeat the overview of all tasks from the preceding chapter:

Always (To):	<i>Introspect carefully and report timely but retrospectively in self-chosen short documentation breaks</i>
Session 1, Task 1 (T1.1):	<i>Warm-up tasks</i>
Session 1, Task 2 (T1.2):	<i>Imagine F_2</i>
Session 1, Task 3 (T1.3):	<i>(Plan to, and then) draw sketches of F_2</i>
Session 2, Task 1 (T2.1):	<i>Imagine F_2</i>
Session 2, Task 2 (T2.2):	<i>Imagine F_3</i>
Session 2, Task 3 (T2.3):	<i>(Plan to, and then) draw sketches of F_3</i>
Session 3, Task 1 (T3.1):	<i>Imagine F_3</i>
Session 3, Task 2 (T3.2):	<i>Imagine F_4</i>
Session 3, Task 3 (T3.3):	<i>(Plan to, and then) draw sketches of F_4</i>
Session 4, Task 1 (T4.1):	<i>Imagine F_4</i>
Session 4, Task 2 (T4.2):	<i>Imagine F_5</i>
Session 4, Task 3 (T4.3):	<i>(Plan to, and then) draw sketches of F_5</i>

Table 12: An overview of all tasks in the main sessions.

3.3.4 Before T1.3: Initial setup – Associations, actions and motor imagery

This section deals with how subjects understand and work on our task, before they are confronted with the first sketching task (T1.3), i.e. at this point subjects have finished with the warm-up tasks (T1.1), are instructed to introspect and report (To) and to imagine F_2 for the first time (T1.2).

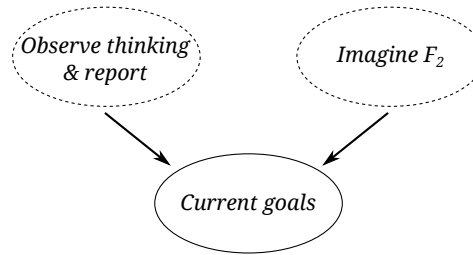


Figure 28: Both, To and T1.2 influence the subject's understanding of what they have to achieve.

Given that we are interested in how people set-up their representations, this is one of the most interesting parts of the study, but since we do not have any data about their thinking before that point in time, also one of the most difficult to study²⁵. However, there are still a few commonalities about which we can report.

When subjects see the first cross-folding task (T1.2), they usually tend to *underestimate its difficulty* and deem it a rather natural, everyday task²⁶. This was sometimes reported explicitly, and otherwise indicated by the short time subjects spend with this task before feeling confident to proceed to the next one (rarely more than a few minutes). This should be at odds with our requirement that the task should initially be complex (TP4). However, even

²⁵ The subject's reports about the preceding warm-up tasks (T1.1) might give us some idea about their general style of thinking and particularly prominent features, but in none of our pilot cases this information would have been enough to know precisely why the subject went on like they did in their first minutes of working on T1.2.

²⁶ This observation is also corroborated informally by numerous experiences with introducing the task to friends and colleagues.

though subjects typically proceeded to T1.3 rather quickly²⁷, they then had to return to T1.2, since they could not fashion the sketches without having to imagine further, such as instructed.

Since this differs largely in comparison to subsequent tasks, this indicates that at the outset, subjects *understood the task differently*. This means, that T1.3 apparently introduces new aspects of the task which the subjects do not consider before: Having been given no specific task beyond “acquaint yourself with imagining the resulting object”, subjects (implicitly) set their own criteria of when they are satisfied with what they are imagining.

At the beginning, most subjects *associate familiar pieces of knowledge*. These might be mental images or analogies, but most prominently they are *actions* related to folding, either embedded in an imagined situation, or referred to abstractly. This means most subjects do not appear to imagine the folded sheet in much detail, but rather they imagine the action of folding it – without paying a lot of attention to what these actions do to the object²⁸.

This might strike us as less odd, when we consider that these imagined actions appear to be based on familiar knowledge and experiences (memory episodes, motor imagery, etc.). In an everyday context, when we need to represent the consequences of an action, we tend to represent those consequences which are *relevant to our respective current goals*. And most people’s prior folding experience consists of folding objects *for reasons other than folding itself*: We fold a letter to fit in an envelope, or our ballot paper to retain the secrecy of our vote – but rarely do we fold in order to appreciate the geometric structure of the sheet that has been folded²⁹. Hence, even though cross-folding is an everyday *activity*, we should expect little prior experience with the *consequences* having cross-folded has on the sheet³⁰.

The criteria they use to decide whether they have fulfilled the goal of “acquainting themselves” might not be much more than thinking of the actions of folding the sheet once, then a second time, and making sure they have done so. Some subjects also started out by including their hands in their imagery, however the majority of reports mentioning hands indicate that the hands are experienced *kinaesthetically*, rather than visually³¹.

Generally, the beginning of a report mostly contains descriptions of associated images, folding actions, accompanied by motor imagery, the contextual background of the actions and sometimes kinaesthetic feelings such as imagined sensations in the fingertips.

S2§1.1	I’m seeing the sheet in front of me and fold it as if I’d close a book	ich sehe das blatt vor mir und falte es als ob ich ein buch schließen würde
--------	--	---

²⁷ This was observed in S2, S5 and S6, i.e. each subject which was ignorant about the exact nature of the test in T1.3.

²⁸ This may seem to contradict the task instruction. Yet, especially for objects which are explicitly shaped by us manipulating them, the notions of an object representation and the actions applied to this object are arguably very closely inter-related (Wiener, 2008). Hence, imagining “only” the actions could be seen as the precursor to an actual object representation.

²⁹ There are activities in which the folded paper is the object of attention, such as Origami. Yet, Origami does not feature iterated cross-folding.

³⁰ The most common property of the object that is mentioned in this very early phase, is that each time we fold, we will halve the size of the sheet. Yet, this fact is rarely of large relevance in later stages.

³¹ Also, actual hand gestures were not uncommon at this stage, but their occurrence was not systematically assessed.

S2§1.2	I imagine how my index finger begins to fold again at the middle of this new construction. (...)	ich stelle mir vor, wie mein zeigefinger inmitten der neuen konstruktion ansetzt um es erneut zu falten. (...)
S2§1.16	I feel in my hand that it wants to rotate the imagined object	ich spüre in meiner hand, dass sie das imaginäre objekt drehen will
S5§1.5	An association to my fax machine appeared to me. On it, one can see a small engraving showing an A4 page which is creased in one corner. Which corner it is, is not graphically present but only abstractly, it is the right one.	Mir kam die Assoziation meines Faxgerätes. Darauf ist eine Einkerbung zu sehen, die ein din a4 blatt zeigt, das an einer Ecke abgeknickt ist. welche ecke ist nicht graphisch präsent sondern nur abstrakt, es [ist] die rechte.

These two examples show excerpts from the beginning of two subjects' protocol of the first session. Both have associations to something that was not contained in the task instruction: S2 to an action (opening and closing a book, rotating it) and S5 to the memory of a visual impression.

Metaphors and motor imagery

Associations like the examples above can be very important, since the appearance of such idiosyncratic images often implicitly *specify parts of the folding procedure*, such as sheet orientation and folding procedure. E.g. the image of opening and closing a book reported by S2 above implies that they made their first fold vertical and towards themselves.

Later on, several subjects employed these initially associated images again in order to quickly re-establish a certain imagined situation. Further examples for early associations we encountered in our pilot studies include folding a newspaper, a body-based metaphor of shoulders and arms moving towards each other, and the inside of the mouth of the Sesame Street's Kermit the Frog³².

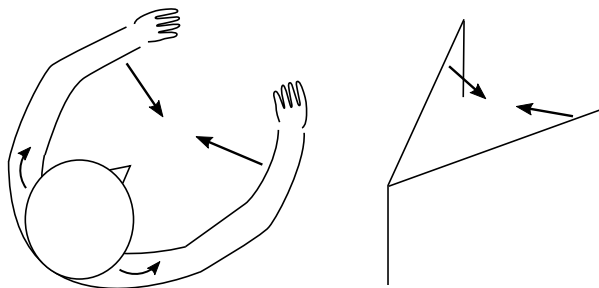


Figure 29: An illustration of the “shoulders & arms” image reported by S3: The symmetrical two-limbed nature of the human torso being a natural metaphor for the folding movement.

Note that many of these images are not static, but many involve motor imagery and are characterised by *movement*, which the subjects identify with

³² This image was only shortly mentioned in S3's report, so we cannot be certain which aspect of the sheet it was meant to express exactly, but note how Kermit's mouth is quite similar to a horizontally folded F_1 and Kermit's talking and erratic head movements correspond well to the actions of folding/unfolding and rotating the sheet.

corresponding transformations of the sheet (folding, rotating etc.). Thus, images like these can constitute something like the first blueprint of a representation of how their imagined folding actions manipulate the object³³. However, such metaphorical images were not reported by all pilot subjects, i.e. not everyone seems to use them for setting up their object representations (or if they do they deem it too unimportant to include in their report). Equally, not every image that appears to a subject is useful or identified as being useful. The image of the paper symbol on the fax machine in the example above, for instance, did not play a further role in S5's protocols.

The question of whether an image is a metaphorical one might be one of degree: While the Kermit-image certainly seems highly idiosyncratic and "creative", closing a book or folding a newspaper is very close to actually folding a sheet and could almost be characterised as a within-domain association.

So, while at this stage subjects associated images – some of which being more metaphorical than others – the commonality might be that *familiar knowledge* is being recruited as blueprint for the folded paper, whereas the "metaphoricity" of some images might possibly serve an additional, mnemonic function.

First specifications

With or without associated images or movements – the specifics of how a subject naturally starts to imagine laying out and folding the sheet tend to persist for the remaining sessions. The earliest details that are specified usually are *sheet orientation* and *folding directions*. It is rare that they are featured explicitly in the report – rather subjects perform the task in a specific manner, such that we can infer from the protocols and drawings that such-and-such *must have been* the direction or sheet orientation they were using³⁴.

A specification which is of little consequence for the first session's tasks may also already occur here: Whether the subject *rotates* the sheet before or after folding, such that it always has the same orientation before folding again or whether they fold in place (cf. Figures 10 and 11). This might lead to very different kinds of errors and patterns of discovery at later points in time (cf. Section 3.3.7).

Rotation must not necessarily be considered as an imagined activity – some subjects just assumed the sheet to be in a certain orientation before folding, and explained this situation by mentioning rotation only when asked to elaborate after the session. This is also an important difference in itself, since only imagining the *result* of rotation can make subjects overlook the rotation's *direction* (which is inconsequential for the first two folds), and hence lead to a situation where they would have had to alternate directions without them knowing about it.

33 The sophistication such an approach may reach, can be seen later in the example of S1's use of "Dracula rising from his coffin", cf. Section 3.3.6.

34 We call such inferred properties of a subject's thinking or behaviour *implicit*, however we do not want to claim that the subjects are necessarily unaware of these properties. Hence, we could equally well call them *tacit*.

Summary

In the beginning, subjects gear up their first representations of the sheet. These are built from whatever the task instruction makes come to mind: Associations to actions, familiar situations or metaphorical images³⁵. These associations might implicitly specify parts of the folding procedure, without a decision between alternatives taking place explicitly. This initial phase typically is very short, yielding little data to analyse. However, more information about these first steps (such as evidence for the subjects' criteria for proceeding) might be inferred to in retrospect, by looking at the expectations and difficulties subjects encounter with T1.3.

3.3.5 T1.3 and beyond: First drawings

The first, very open phase of imagining described above is disrupted when the subject is first asked to fashion drawings (i.e. Session 1, Task 2) and they hence have to try constructing a detailed representation of the F_2 .

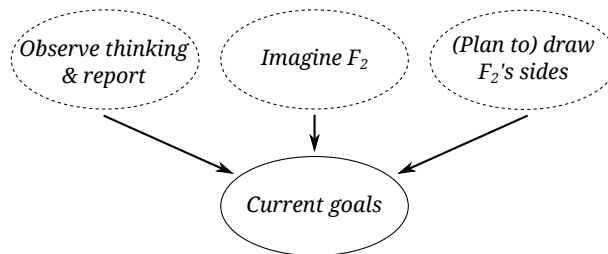


Figure 30: To achieve T1.3 most subjects have to keep working on T1.2, i.e. T1.3 adds to the subject's current idea of what they have to do.

Usually, they discover that they are not able to do so right away, even though they were confident they could imagine the object well before³⁶. So as a direct consequence of T1.3, they are faced with two new sub-tasks:

1. Find out how all sides of F_2 look and remember them long enough to draw them,
2. Find out the position and orientation of each side and remember them long enough to note them under the correct drawing

There are more than one strategies to "remember" (in the broadest sense) how the sides look, trying to memorise a mental image in detail being only one of them. Another one would be to remember how the side is being constructed (i.e. how folding the sheet affects the edges on this side), or in later stages when this knowledge is available: Describing it in terms of composed regular figures (Vs, nestings etc.) and reproducing this description later on. This strategy choice varied widely between subjects and between sessions, however the explicit occurrence of this as a "problem" on its own appears only later, when the sides get more complicated.

³⁵ With *metaphorical image* we mean a mental image which takes form of something from outside of the paper-folding domain, but which is taken to mean some within-domain state of affairs or action.

³⁶ Remember, subjects choose to proceed to the next task by themselves, asked to do so when they feel confident to be subjected to a test of their understanding.

However, the main problem at this stage is the first part of the first sub-task above: *Find out how all the sides look*.

Getting to know the sheet

Subjects engaged in diverse imaginary activities in order to acquire detailed knowledge of the folded sheet. Most of these correspond to actions one could also apply to a physical sheet, such as (a) repeatedly unfolding and folding it again, (b) rotating it in various directions, or (c) changing their perspective on it. These actions are typically executed *paying attention to specific parts of the sheet* only, such as single sides or single edges. This implies either that subjects are not able to rotate the whole folded sheet as one object, or that while they are able to do so, they still can only pay attention to a part of it³⁷.

One action which several subjects imagined was to stroke along the edges of the folded sheet with their fingers, an activity we called *finger tracing*. An interesting aspect of this action is, that while no subject phrased it like this in their report, as an activity following the sheet's edges with one's fingers is very similar to drawing with a pen (or using one's finger as a pen when "drawing in the air". In other words, this activity might be interpreted as an action mediating between the imagined sheet and the strokes one would have to make when drawing the 2D-projected sketches.

First discoveries: The V and its relatives

When investigating F_2 in detail, subjects are bound to discover their versions of the primitive figures (cf. Figure 13): The two rectangles³⁸, the crease (I, stroke, ...), the V (varyingly called "U", "C", "greater-than sign" or "arc")³⁹, the DV and the NV.

Not all of this structure has to be noticed and named by subjects in the first session (although rectangles, crease and the single V usually are), but in any case, it is important to remember that at this point they are names for concrete, individual sides of F_2 , and subjects usually do not anticipate that these structures will re-occur in later folds and could be generalised as central features of cross-folds.

Perspective and canonical tags

Our second task asks subjects not only to fashion sketches of all sides of the folded sheet, but also to annotate which sketch represents which side (the second sub-task we introduced above). There was a lot of variation in how difficult this appeared to be for our pilot subjects, indicating that while some of them must have had explicitly considered perspective when imagining the sheet before (and made it "canonical" to some extent), others

³⁷ The ability of "how much" of the sheet was possible for a subject to imagine at one time was highly dependent on the knowledge they had available to structure their imaginations ("chunks", if you will). For some subjects, finding a way to get an impression of the whole folded sheet in most of its details became a task of its own. We discuss some of their approaches in Section 3.3.6.

³⁸ Typically, they are ignored very quickly and without further mention.

³⁹ Note that up to now every subject has naturally imagined and drawn the sheet's sides in a such a "slightly ajar" manner.

seemed to have imagined the sheet more “abstractly” without considering a space with named directions. This might have to do with how strongly a subject thinks that the *orientation* of the sheet matters. Whereas, if all that is important is the geometric *configuration of edges* on the sheet’s side, then orientation can be ignored – since the relation *between* all edges remains equal and only the sheet as a whole is differently oriented in space.

Independently of this issue, however: Once a subject has annotated directions to their drawing⁴⁰, these directions (and hence the sheet’s orientation) tend to be used again for the following sessions.

Note also that for many subjects there appears to be a difference between their tags for the sides of the canonically oriented sheet, and other current orientations as they might consider when imagining moving around or rotating the sheet, i.e. subjects distinguish between their egocentric and the sheet’s *frames of reference*, leading to statements like “I’m looking at the left side as if I was facing it.”.

Depending on whether the subject uses a folding procedure with or without rotation the figures which will be assigned these directions might reveal different regularities in future sessions (cf. Section 3.3.7).

The actual drawings

At the end of Session 1, subjects produce their first sketches. This not only provides an opportunity for the investigator to make sure the task was understood correctly⁴¹, but the sketches and observations about how they were produced are also a valuable additional data source:

Is the sketches’ relative orientation corresponding to how it would be on the sheet? If so, they could provide evidence about how the subject imagines the sheet (especially in the first session the quality of the descriptions in the protocol often does not suffice alone to resolve such questions)⁴². The order in which the sides are being drawn could also indicate in which order the subject tends to traverse the sheet mentally (if there is a stable order at all). And finally, we might notice whether the subject tries to model other properties of the sheet (e.g. aspect ratio).

Summary

When asked to draw F_2 ’s sides (T1.3), the subject’s attention is shifted to imagining F_2 in more detail. The necessity of this shift indicates that our task indeed requires subjects to change their representations (TP1). Exploration of the sheet through repeated un-/folding, rotating and perspective-changes leads them to discover first concrete forms such as the V (variably named). Imagined actions like finger tracing might allow subjects to mediate between the imagined sheet and the imagined action of drawing the 2D sketches.

⁴⁰ And once these annotations make sense – sometimes the annotated directions are not coherent with *any* orientation, indicating that the subject is likely to only have constructed the individual side figures, but did not coordinate them with a specific state of the whole sheet.

⁴¹ Our first pilot subject, for instance, misunderstood the task to be asking for a *parallel* fold, i.e. one in which the subsequent creases are not orthogonal but on the same axis. This was only noticed and clarified when the investigator saw his sketches.

⁴² Note however that we should not to assume a perfect correspondence between a sketch and a subject’s mental imagery – there are many additional processes involved in fashioning a physical drawing.

Finally, the subject's perspective on the sheet and the sheet's orientation are likely to become canonical.

3.3.6 Dealing with complexity: First abstractions

With an increasingly number of folds, finding out how the sheet's sides look by simply attending to them in imagery becomes more and more difficult. This means that with F_3 or F_4 , subjects have to find different ways to solve their tasks. These "different ways" generally originated out of the subjects – explicitly or implicitly – gaining more knowledge about the task domain. In this section, we describe the first abstractions that took place and regularities (about the task domain) that were discovered. Specifically, we discuss rule-like regularities, metaphorical images, and abstractions implied by the drawing task.

At the latest at this stage (Sessions 2 and 3), our pilot subjects' thinking diverged widely. At their broadest, these differences are apparent in the subjects putting varying emphasis on different sub-tasks:

For some, preparing to draw the sketches had essentially become the main task and imagining the fold was only a sub-ordinate task in that to accomplish the main task, imagining the fold can be helpful. For others, it was the other way around, and imagining the fold was seen as the main task, which could be helped along by thinking about the drawings. And finally, one subject (S5) took care to keep the two tasks separate, as indicated in Figure 30. These different possibilities of comprehending our task also demonstrate that our task provides subjects only with a vague goal (TP5).

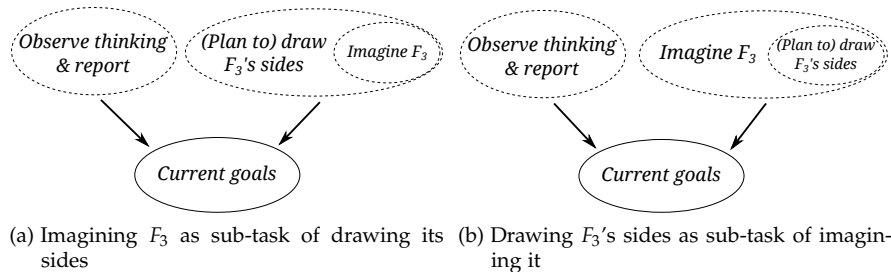


Figure 31: Two differently evolved task comprehensions.

In addition to the "official" tasks and sub-tasks directly derived from them (such as finding out how the sheet's sides are looking, where they are etc.) subjects sometimes also set themselves tasks which are less directly related to fulfilling the main tasks. Often they occur as part of working on the main task, and then they are being followed up on as a curiosity. For instance, these are questions like *how many layers* of sheet there are from a certain side (or how many are visible), or the idea (which several subjects had) to *try seeing the unfolded sheet in the folded sheet*, i.e. to try imagining the whole object not as complex configuration of edges, but as a single, multiply folded surface. While for some of these questions it can be difficult to see their relation to the main task, the exploration of the cross-folding domain and possible regularities discovered by doing so, can have a large influence on the subject's thinking.

Furthermore, the discovery of regularities seems not only driven by having to deal with the complexity of higher folds, but also facilitated by simply having repeatedly imagined F_2 , F_3 at this stage, leading many subjects to start noticing regularities both between folds and within the same fold. Early *between-fold regularities* were for instance, that each fold appears to have structures like a V, a DV etc., or that the I is folded into a V, whereas early *within-fold regularities* would be that the DV is opposite to the I, or that the V is perpendicularly adjacent to it, etc.

Interestingly, not all of these regularities were always attributed the same generality – while it is immediately apparent that *every* fold (including ones the subject has not imagined yet) has an I (a crease), noticing DV_2 and DV_3 did not necessarily imply that having a DV was automatically postulated as a general feature of cross-folds. While there might be many factors involved, the willingness to make such an inductive inference often seems to depend on how easy the regularity is to derive from intuitively imagining the folding procedure.

Different kinds of regularities

When constructing F_3 there are many opportunities to discover structures similar to the ones in F_2 , even though subjects do not necessarily explicate these observations. Sometimes, a discovered regularity might only implicitly modify a subject's behaviour. For instance, most subjects left out drawing the two rectangular sides from F_3 onwards (just as we did in Figure 13) without further mention (indicating their knowledge about the constancy of the rectangles appearance).

Another regularity which became apparent to subjects relatively early (judging from how they produced their drawings) was that in each fold on one side there is a single stroke (the crease). When imagining F_3 , we might notice that NV_3 is clearly a nested structure, just as NV_2 . But when constructing it mentally, it might make more sense to think about it as a creased DV_2 , without considering that it is nested at all.



Figure 32: Two different descriptions of NV_3 .

Both descriptions of NV_3 in Figure 32 are already capturing a lot of domain knowledge. Yet, regularities postulated⁴³ by subjects can also be more concerned with the “surface appearance” of a figure. For instance, instead of saying that “*every fold, the preceding fold’s crease is folded into a V*” one might also note that “*every fold, on one side there is a single V*”. Both rules describe

⁴³ The word “postulate” is not meant to imply that subjects put their ideas forward as general statements, merely to say they *assume* that some relation they notice holds, either implicitly or explicitly.

the same figure, but the first one incorporates much more knowledge about the task domain⁴⁴.

These two kinds of rules can be understood as pertaining to one of two different styles of thinking, in which people predominantly notice either constructive relationships like in Figure 32(a) or structural relationships like in Figure 32(b). These were described by Schwank (1999) and called *functional* and *predicative thinking*. While our task mostly provokes functional thinking (through its constructive nature), predicative thinking also plays a role (varying between and within subjects).

The advantage of noticing and postulating regularities with regard to the task is that they can be used either to help along and scaffold imagery (e.g. help not getting lost while paying attention to different sides, because we have declarative knowledge about which side is where), or even to use them *instead of* detailed imagery. This lack of imagery can be inferred from the occurrence of errors resulting from the wrong application of rule-like regularities which would have been easy to catch in imagery (such as leading to a physically impossible state of affairs). Section ?? will discuss an example of this kind of error from S6's data.

Metaphorical images

Discovering rule-like regularities was not the only way subjects tried to find structure in the more complex folds. In many cases, subjects noticed a task-related idea spontaneously embodied in the form of a *metaphorical image*. Such images served many different functions: Some described a specific fold in one elegant image, some describe general relations between specific parts of folds or of the subject's folding procedure, while others were used communicatively to describe a mental image more closely for the report⁴⁵. Since these metaphorical images had numerous functions (some of them more than one at the same time), in the following we will present several examples from our pilot studies.

Especially in phases when the subjects were concentrating on the visual imagery of the folded sheet, they came up with metaphorical images of "how the sheet looked like". Figure 33 shows illustrations of two of such metaphorical images (reported by S1 and S3 respectively). Superficially, they are both images based on structural similarities to F_2 , depicting it in a specific orientation, but there are large differences in the details: First, the image of Dracula with his cape actually *captures the whole geometry* of F_2 (the cape), whereas the image of two-peaked mountain massif depicts the two Vs of the DV and their orientation, and suggests that they are connected through the mountain crest bending around in the background, but without depicting any detailed structure. When we look at how the mountain image was used by S3, we find that in its first occurrence it appeared as passing remark

⁴⁴ Telling the two rules in the example above apart is quite easy, since they were stated so clearly and explicitly. But generally, in interpreting a subject's data, it can be very difficult to find out exactly how general or specific something was understood – sometimes a specific name can be used to express a general state of affairs, and sometimes the general term is used, yet only the specific instance was thought of (*synecdoche* or other kinds of *metonymy*).

⁴⁵ Note that what we have said about the interpretation of rule-like regularities above applies to metaphorical images even more strongly: It can be very hard to tell how much of what the verbal description of such an image entails was actually consciously present when the subject thought of it, and if so to what extent. Especially since the mnemonic nature of a useful metaphor usually means that subjects keep returning to and refining them.

meant to *express the visual impression* of the “massivity” of looking up at DV_2 , but when the image reappeared later on several times, it was mostly used to *re-establish sheet orientation*.

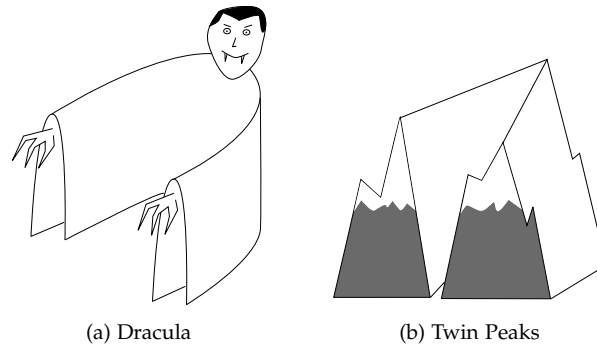


Figure 33: Illustrations of two metaphorical images for F_2 .

Similarly, the Dracula image had more uses than to depict F_2 . Actually, it appeared in S_1 's thinking about his folding procedure when trying to construct F_3 . When folding the sheet horizontally and then rotating it 90° counter-clockwise, as was S_1 's canonical procedure, the image of “a vampire rising from his coffin” appeared to him. Since the cape only describes the structure of F_2 , when trying to see Dracula as F_3 , the spread fingers were taken as indications that more details have to be worked out at these points.

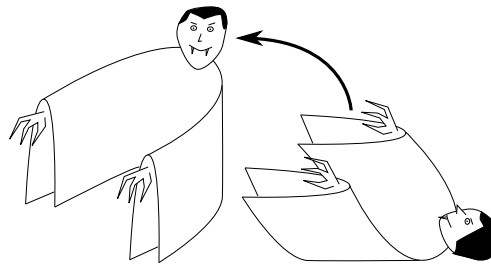


Figure 34: Illustration of a further aspect of the Dracula image: The sheet being rotated after folding is the vampire rising from his coffin.

Not all metaphorical images which appear to subjects are drawn from remote knowledge domains. Figure 35 shows another image that appeared to S_1 in a later session, when he went successively through his folding procedure, and noticed that F_1 can be taken to be an “extruded” V – one V which has been elongated into the third dimension. This allowed in turn, to see F_2 as a single, folded surface, consisting of this extruded V which has been folded in the middle.

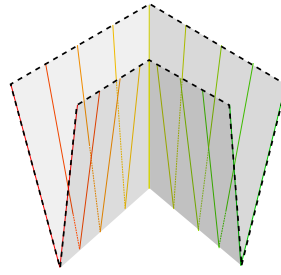


Figure 35: Illustration of another image-like interpretation of F_2 as consisting only of “extruded” copies of Vs. The coloured Vs are there to signify this extrusion, not to indicate specific structures. The dashed black line indicates the original edges of the unfolded sheet.

Since the drawing task suggests a partition of the folded sheet into sides, some subjects soon started to solely focus on the sides as separate entities and thus often confused sides with each other (since they did not know where they belonged). Metaphorical images providing such a “wholistic” perspective on the sheet as the examples above might play an important role in preserving and/or attaining overview of the sheet’s geometry, although there were subjects who managed solving the task without reporting any such images (S6).

A final example shows a metaphorical image which neither depicted the whole sheet, nor a single side, but which put two sides *into relation*. In their first session, S4 noticed a figuration built from F_2 ’s crease and its (perpendicularly adjacent) V, the “crow’s foot” (Figure 36). Note that when this image first appeared, S4 was not aware of its possible generality and usefulness.

S4§1.21	The old crease is on the bottom and has become a single V; an attempt to quickly relate it with the new crease yields a constellation of three “lines”, the “lying” V made from the old crease, and proceeding upwards from its closed end an perpendicular line – I associate this with a crow’s foot.	der alte falz ist unten und zu einem einfachen V geworden; ein schneller versuch ihn mit dem neuen falz zusammenzudenken gibt mir eine konstellation von drei „strichen“, das „liegende“ V aus dem alten falz und von seinem „falzpunkt“ ausgehend ein senkrechter strich nach oben – dieses gebilde assoziiert einen hahnenfuß.
---------	---	--

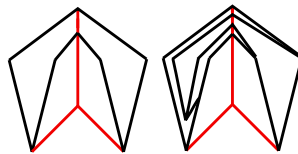


Figure 36: An illustration of the “crow’s foot”-image in F_2 and F_3 .

In later sessions, S4’s memory of the crow’s foot was among other things used to imagine F_2 in a different orientation (cf. Figure 37), before it was “re-discovered” in another position⁴⁶ in F_3 :

⁴⁶ S4 used a non-rotating folding procedure.

S4§3.23	To my surprise, the new line at the top and the roof-V at the right form a new crow's foot.	der neue strich oben und das dach-v zur rechten ergeben relativ überraschend einen neuen hahnenfuß.
---------	---	---

The crow's foot is interesting because it is very simple and – in its generalised form – it embodies several important properties of the cross-folding domain. Namely that each fold has a crease and a V, that they always are adjacent and that they are orthogonal to each other. Finally, being a figuration of all of the closed sides⁴⁷ of the sheet, it can serve as an easy to handle representation of the whole sheet's orientation in space.

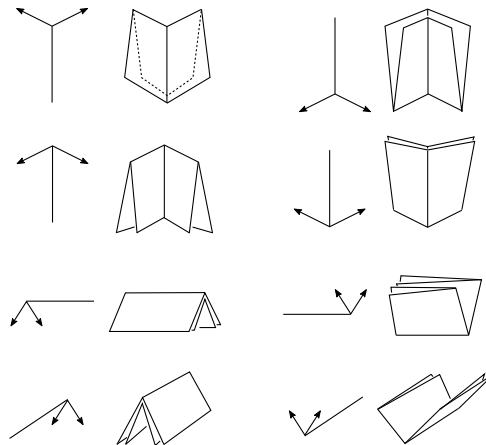


Figure 37: Eight of the many possible orientations of F_2 . The crow's foot allows to easily depict the sheet's orientation without having to consider the more complicated, open sides.

Closing our discussion of metaphorical images and related phenomena, we might say that even though the remote and varied fields of knowledge they are sourced from may make them appear as an disordered collection of random occurrences, there are several commonalities we can note: (a) the first occurrence of a metaphorical image seems to be spontaneous (judging from our data), (b) most images mentioned in the reports have the tendency to re-occur later, (c) they fulfil mostly *structuring* and *mnemonic functions*.

Abstractions implied by the drawing task

We already mentioned that due to the complexity of F_3 it is unlikely that subjects manage to imagine it with all its detailed structure present concurrently, and thus they are "forced" to think more abstract. But it is not only the complexity of higher folds which suggests thinking more abstractly: The task of drawing "schematic 2D sketches" of the sides itself suggests that certain "realistic" properties of the sheet are not important. Hence, the subjects are less and less likely to spend time on imagining such aspects of the sheet, e.g. aspect ratio, colour, texture or thickness of the fold. If one of these occurred nevertheless, subjects tended to think of them as distractions:

⁴⁷ "Closed" meaning we are looking at the creases as if they were folded away from us, obscuring the "insides" of the fold.

S4§2.20	a second attempt yields a tidy second crease, attention is focussed on making it 'edge-y', the imagery is quite vivid, but not much fold-form was kept in mind.	ein zweiter ansatz macht den zweiten knick sauber, das augenmerk liegt dabei darauf ihn 'kantig' hinzubekommen, die vorstellung war relativ stark bildlich, aber es wurde nicht viel falt-form mitgedacht
S4§2.21	(i'm quickly ruminating that this aspect is irrelevant and will only distract me from the sheet's form, if i follow up on it)	(ich räsioniere kurz, dass das ein irrelevanter aspekt ist, der mich nur von der blattform ablenken wird, wenn ich ihm nachgehe)

Another such idea which occurred to several subjects at some point was that repeated folding of a piece of paper becomes physically impossible after relatively few folds (cf. Section 3.2.3). Other such realistic concerns that we found in the protocols were worries about getting the crease straight or that the sheet becomes overly small after a few folds, leading to imagining stretching it.

Even though these "distractions" do not form a part of the idealised task domain we described in Section 3.2, thinking about them might still turn out functional for a subject. They might for instance be a source domain for analogies, play a role in strategy change or set a new curiosity-driven sub-task. For instance, thinking about the thickness of a physical fold might lead to the question of finding out how many layers of paper are hidden within such a closed side and wanting to explain how they come to be there.

Summary

After having imagined several folds in a short time frame, the subjects' attention shifts from the concretely imagined sheet to more abstract pieces of knowledge – whether these might be metaphors, images, regularities in surface appearance, regularities regarding domain principles or implicit knowledge we observe in the behaviour. This shift is mainly brought about by the longer experience with the task, the increased complexity of $F_{\geq 3}$, and the schematic nature of the sketches subjects are instructed to draw. The complexity subjects are faced with, and their discovery of regularities which allow them to deal with the task nevertheless demonstrates that our task indeed exhibits high, but reducible complexity (TP4). Additionally, the multitude of possible kinds of representations involved at this stage further demonstrates that our task requires representational change (TP1).

3.3.7 Systematising sketch generation

In the preceding sections we concentrated on the increasing difficulty of *imagining* the sheet and its sides. Here, we concentrate on the difficulty of planning to draw and then actually *drawing the sketches*. While F_2 and maybe F_3 are still simple enough to be drawn unplanned – provided the subject knows how they should look – with the increasing complexity of higher folds *planning* the drawings (e.g. how to sub-divide the drawing, in which order to draw, starting from which end etc.) becomes a major problem by itself.

Drawing strategies may differ widely in the amount of mental preparation, as well as in how much manual dexterity they require. In fact, those two

measures have a roughly inverse relationship: Strategies that need more planning and use more complex cognitive structures and abstractions tend to be easier to execute manually on paper and vice versa (trading off cognitive against manual complexity).

Three drawing strategies

The cognitively simplest drawing strategy would be to just draw each individual sub-figure with one continuous stroke. To do so, we “only” need a description of each sub-figure in enough detail. So, (given that we already are confident in our use of the primitive forms) in the case of NV_5^I as seen in Figure 38 below maybe: “There is one large doubly-folded V – inside of which there is another large doubly-folded V, and nested within these there is one singly-folded V”:

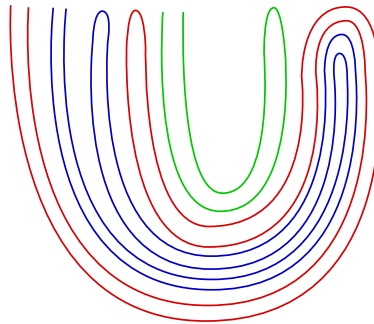


Figure 38: Illustration of a procedure for systematically drawing a sketch (here of NV_5^I): Draw each individual sub-figures in one stroke (in red, blue, green).

While it sounds pretty straight-forward, this strategy very soon begins to show weaknesses: As the figures get more complicated, a lot of care has to be taken such that all sub-figures fit beside or inside of each other⁴⁸. This can lead to “cramming” or even having to start all over again if the subject notices that they have misjudged the space needed for each sub-figure⁴⁹.

Experiencing, and from then on anticipating this difficulty may lead to changes in the drawing strategy: For instance, in his last session, S6 came up with the idea to first draw the correctly aligned and generously spaced “tips” of each sub-figure, and only then proceed to draw the folded edges (Figure 39). As opposed to the naïve strategy, this requires knowledge about how many open and closed tips there will be. This could either be abstracted from direct knowledge about NV_5^I or, using one’s understanding about how the NV-sides of a fold are generated, from remembering the tips of DV_4^I – which are the same, cf. Figure 32(a).

⁴⁸ In fact, making a freehand drawing of NV_5 without mistakes, such as shown in Figure 38, is already virtually impossible without a lot of practice.

⁴⁹ Even though subjects are instructed to only start drawing when they have planned what they want to draw *exactly*, considerations like spacing often occur on the spot (which again serves as reminder that these physical drawings must not be taken for direct mappings of a subject’s mental imagery).

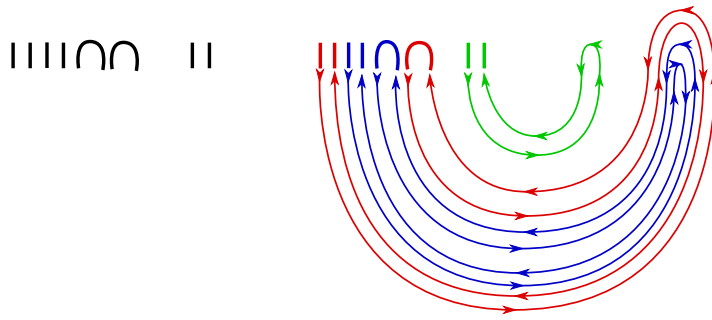


Figure 39: Illustration of a procedure for systematically generating a sketch (here of NV_5^I): First draw ends from one open side of the new fold (left), and then draw individual sub-figures (right; in red, blue, green).

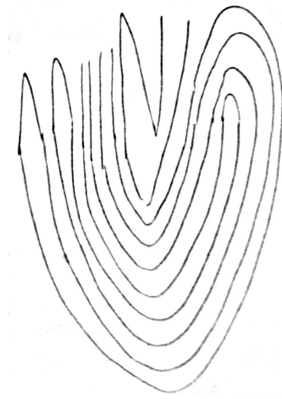


Figure 40: A (partly incorrect) drawing of NV_5^I fashioned by S6, employing the drawing strategy explained above (image mirrored to allow comparison with Figures above).

The easiest way of drawing the sketches, is to not only prepare the figure by drawing the tips of sub-figures but also the nested V-edges on the other side of the fold (cf. Figure 41). All that needs to be done then is to join these tips and nested Vs with arcs, which can all be drawn independently (thus requiring far less manual dexterity than the other strategies, where we have to draw arcs back and forth). To be able to do so however, fundamentally hinges on the subject being able to predict the top ends of all sub-figures. In our example: For the other strategies it sufficed to know that there are two nested and doubly folded Vs (red and blue), but for this strategy we have to know that jointly, these folded Vs produce a nesting of four Vs at the rightmost side of the figure.

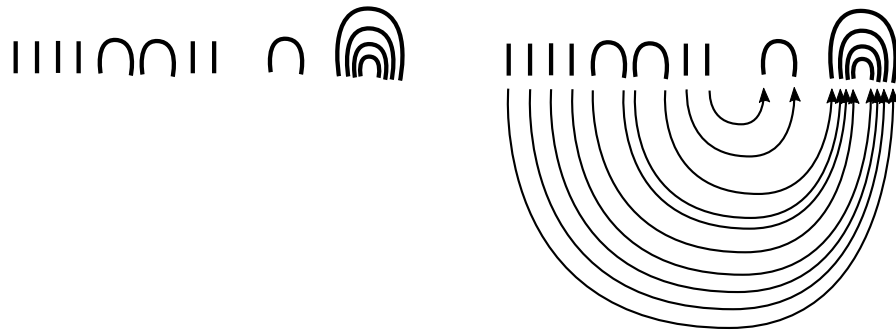


Figure 41: Illustration of an easier procedure for systematically generating a sketch (here of NV_5^I): First draw all ends (left), and then join with (more or less) concentric arcs (right).

Possible discoveries

After already having discovered the first patterns, the reasoning necessary for systematising one's drawing strategies – namely having a highly structured way of describing side (or sketch) composition in terms of Vs, nestings etc. – should lead subjects to further discoveries. Since more advanced drawing strategies only started to appear towards the end of our last session, we do not have much data from our pilots, indicating how they proceeded from these strategies. But this structured knowledge might be an essential contribution to developing a generalised procedure of generating the sides of F_n , and especially the last drawing strategy we described above could ultimately form the basis of a *syntactic procedure*, such as described in Sections 3.2.9–3.2.11.

Dissociation of sketch composition and folding knowledge

The rather abstract knowledge about sketch composition can not only be advantageous for a subject's reasoning, but it can also lead to specific errors. Namely, a subject might start to describe a fold's sides and its sketches solely in terms of figure composition and *stop checking in imagery* whether they actually are folding correctly. Thus, especially in the early phases of discovering sketch composition knowledge, it might actually be used *instead* of the folding knowledge, as opposed to being coordinated with it. This could ultimately lead to a dissociation of task-relevant knowledge into two separate "domains"⁵⁰.

Deriving how to draw a sketch purely from their sketch composition knowledge often lead to errors which subjects would have noticed easily when imagining the fold. Figure 42 shows a sketch where S3 drawn a sketch of DV_3 which shows three parallel adjacent sub-figures – something that cannot happen in cross-folding⁵¹.

⁵⁰ This interpretation is reinforced by several subjects reporting on distinct visual imagery of the sheet on one side, and of of lines or pen strokes on the other – a distinction at least one subject (S5) upheld even when imagining the "same" object close in time.

⁵¹ Making mistakes in imagery also lead to erroneous sketches, yet interestingly, these mostly lead to sketches which depicted wrong, but possible sides, whereas the examples here show sides which are impossible to achieve by cross-folding.



Figure 42: An incorrect drawing of DV_3 fashioned by S_3 : Three Vs next to each other.

As a further example, consider the following sketch drawn by S_5 in which she wanted to depict NV_3 : Her sub-problem at that time was that while she knew that figures were composed of “Us” – which could be either alone, next to one another, or within each other – and that the NV_3 had to be a nested figure, she did not know *what* exactly was supposed to be embedded in it. Thinking that it is always a figure from a preceding fold (which is not wrong, but they have to be folded further), she apparently nested DV_2 within NV_2 .

Yet, imagining to fold F_2 into F_3 could have told her that these sides cannot end up nested like that – so it seems her drawing was made with the help of her sketch composition knowledge which, at this point, was not sufficiently constrained by her folding knowledge.



Figure 43: An incorrect drawing of NV_3 fashioned by S_5 : A double V nested within two nested Vs.

Summary

Since with increasing fold complexity, not only imagining the fold, but also planning and drawing the sketches becomes more difficult, developing strategies to plan the sketches becomes a vital part of our task in later sessions. Strategies may differ with regard to how much knowledge of the sides they require and regarding the difficulty of executing them manually. The knowledge gained by acquiring an advanced sketch composition strategy may contribute to developing a generalised recursive procedure and form a precursor of syntactic notations, but this was outside of the scope of our pilot sessions. What could be observed in the pilots however, was possible disadvantages of undertaking sketch composition without being constrained by imagining the fold. While errors in fold imagery typically lead to sketches which were erroneous, but still physically possible, sketches

drawn on the basis of faulty sketch composition knowledge often depicted physically impossible situations.

3.3.8 Beginnings of a recursive procedure

In Section 3.2.8, we described a general recursive procedure to generate a certain cross-fold. Ignoring the simpler sides, the recursive rules for the DV and NV were:

- (I) NV_n : fold DV_{n-1}
- (II) DV_n : put V_{n-1} and NV_{n-1} next to each other

While none of our pilot subjects developed such a truly general procedure within the duration of the main sessions, most of them developed procedures which could be interpreted as *precursors* of such a solution. Given our task and the way we present its sub-tasks, the idea to reduce the current sub-task to the sub-task that had been solved before appears quite natural. But while it appears obvious to solve our task recursively, and to construct a fold out of the preceding ones, the extent to which our pilot subjects noticed and generalised this recursive nature of their own procedure varied greatly. The pilot subject who developed a strategy very similar to the recursive one, was S6, and in this Section we will use his case as an example.

Early in Session 2, when first having to imagine F_3 (T2.2), S6 mentioned the idea to generate F_3 's sides, proceeding from the sketches of F_2 he had made yesterday and which he was thinking of when working on T2.1:

S6§2.17	some of the sketches I could produce, I think, if I just think forward from yesterday's sketches. with others I have difficulties	einige skizzen könnte ich - glaube ich - anfertigen, in dem ich einfach die skizzen von gestern weiterdenke. bei anderen habe ich so meine schwierigkeiten
---------	---	--

So already at this stage, he thought of imagining the fold in terms of its sketches, and of producing them by manipulating the one's of the preceding fold, side by side (or rather: sketch by sketch). Based on this idea, he mentally reconstructed his sketches of F_2 and imagined *folding these sketches 2-dimensionally*⁵², leading to his first sketch of F_3 :

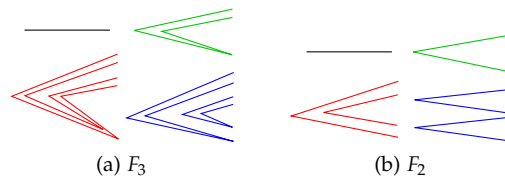


Figure 44: Colourised reproduction of S6's first sketches of F_3 (left) and the sketches of F_2 he constructed them from.

So the (tacit) rule he was following was an overly general version of (I): For all sides except the I:

- (I_{S6}) $sketch_n$: fold $sketch_{n-1}$

⁵² Except for the two rectangular sides and the I, which he already knew always look the same.

Only when he had difficulties in drawing the last sketch of a creased NV_2 (the red one in Figure 44a), he returned to *imagining the folded sheet*, where he soon noticed that his sketches of F_3 were faulty.

At the end of session 2, he had worked out from imagery that *only two sides* are generated like this (the NV and the V), and that two other sketches do not have to be folded, but *stacked*, bringing him closer to a correct version of rules (I) and (II).

(I'_{S6}) for some sides: $sketch_n$: fold $sketch_{n-1}$

(II_{S6}) for other side(s): $sketch_n$: stack $sketch_{n-1}$ and opposite $sketch_{n-1}$

However, the propensity to manipulate mental sketches instead of imagining the sheet lead to another interesting error: When imagining the sheet, S6's folding procedure included rotation, such that before folding the sheet would always be in the same canonical orientation⁵³. But when generating the sketches of F_3 *directly from* those of F_2 , this rotation step was skipped, making him fold & stack the wrong two sides):

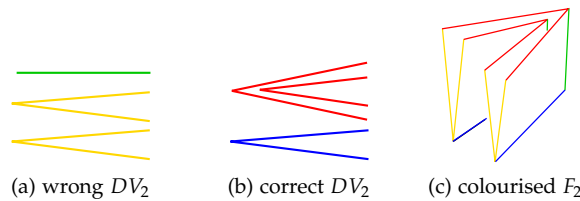


Figure 45: Illustrations of S6's stacking method of constructing the DV, forgetting to take rotation into account. Colourised F_2 included for comparison.

After having corrected this error through a long episode of returning to imagining the fold, S6's fold & stack strategy remained fundamentally the same for the two remaining sessions. However, since these two rules do not really specify how this folding and stacking should work, following the strategy still was fraught with difficulties. For instance, as opposed to F_2 , the sketches of F_3 were too complex to simply remember and manipulate them, leading to many losses and subsequent reconstructions of F_3 from F_2 when trying to imagine F_4 ⁵⁴.

As we mentioned at the outset of this section, other pilot subjects did not get so far in developing a quasi-recursive strategy for generating the sides. Instead they spent considerably more time with detailed sheet imagery as well as reporting. The interesting points in S6's strategy development were that his early focus on avoiding imagery lead to a very efficient, and in the last session almost syntactical approach, but it lead to many errors which could have been avoided if these descriptions would have been *coordinated with imagery* of the sheet and the folding procedure. The report of S6 developing and refining his strategy over the course of several sessions and the necessity of keeping this strategy coordinated between different ways

⁵³ This went so far that instead of identifying the sides of the sheet by their properties (double V, nested V etc), he only talked about the frontal side, the left side, etc.

⁵⁴ This problem of limited capacity indicates that even though the general recursive solution is simple to explain, its execution can quickly become very cumbersome. While computers can avoid re-computing intermediate results of recursive function calls by holding them in memory – a technique known as *memoisation* and likened to “learning from experience” in the original publication (Michie, 1968, p.19) – the limited capacity of human working memory puts strong limits on following such a strategy naïvely.

of representing the fold (sheet imagery and sketch manipulation), demonstrates that the representational changes afforded by our task take place over a longer stretch of time (TP2).

3.3.9 *Conclusion*

This concludes our semi-chronological report of observations about representational-level structures and learning trajectories drawn from the data of our pilot subjects.

We noted how subjects “commandeer” familiar knowledge, memory episodes and situations as blueprints for beginning our task, and how their initial focus lies mainly on the folding procedure and motor imagery, more than on the geometry of the resulting fold. Then, when asked to fashion their first sketches, we observed that the subjects’ focus shifted more towards the object, possibly leading them to discover the first primitive forms (I, V etc) and introducing a canonical orientation for folds. We noted that with more experienced and driven by the complexity of the higher folds, subjects both (a) simplify their imagery by leaving out details deemed to be unnecessary, such as aspect ratio and paper size, and (b) add to their imagery by discovering rule-like regularities and metaphorical images which embody aspects of the sheet or the folding procedure. We saw that the complexity of the higher folds does not only influence subjects’ imagery, but also that their strategies for planning and executing their drawings has to adapt and become more systematic to avoid making errors. We assumed this sketch drawing knowledge could be used to develop a more syntactic comprehension of a fold’s sides, but could not tell from our data, since systematic sketching knowledge – if at all – developed only in the last session of our study. What we could witness in our pilots however, was strong dissociations of folding and sketching knowledge which lead several subjects to fashion drawings which their current folding knowledge could have identified as wrong. And finally, we followed the case of one subject developing and refining a recursive strategy for generating the sheet’s sides, being successively forced by sheet imagery to introduce more constraints into this procedure.

Before we continue to Section 3.4, in which we discuss the possibility of formulating representational-level problem spaces in which all these activities could take place, we prepare for this discussion with an overview of sub-goals which were either formulated by subjects or which we inferred from the subjects’ behaviour as possibly being their current intent.

3.3.10 *Overview of some possible sub-tasks*

This overview only contains some of the tasks and sub-tasks that were discussed in the observations above. Given the nature of our task, it should be obvious that many more possible sub-tasks, goals and strategies could possibly occur, and we would expect that conducting our study with further subjects would lengthen this list considerably.

- Main tasks: Observe & report, Imagine F_n , (plan to) and draw sketches
- Common sub-tasks:

- Find out how sides look
- Find out where sides are
- Standardise folding procedure
- Introduce canonical tags
- Find way to generate sketches more systematically:
 - * Identify sub-figures of each side
 - * Partition sub-figures into simpler entities
- Implicit sub-tasks, brought about by working on the above:
 - Find way to save mental effort
 - * Substitute imagery with verbal-syntactic thinking
 - * Find within-fold regularities
 - * Find between-fold regularities
 - * Find metaphoric images
 - * Memorise sketches
 - Align sketch and sheet imagery
 - Read sketches from sheet imagery of a side
- Curiosity tasks which allow exploration
 - Find out how many layers can be seen from some perspective
 - Find a way to see the unfolded sheet in the folded one
 - Find figurations of edges spanning multiple sides

3.4 BACK TO THEORY: POSSIBLE REPRESENTATIONAL-LEVEL PROBLEM SPACES

In this section, we discuss how we could describe activities such as the ones from our pilots we described above in terms of problem spaces. Given that we did not discuss one of our pilot cases in detail, but rather presented a broad overview sourced from all of our pilot subjects, we cannot introduce individually fitted problem spaces here. Rather, we will present a coarse selection of several problem spaces which might serve as “templates” for the definition of individual spaces, each describing aspects of the subject’s problem-solving activities.

The first question when trying to find problem spaces to describe our subjects’ activities, is *how many problem spaces* we should assume. To approach this question, we start with a look at the criteria put forth by Schunn & Klahr (1996) which we already quoted in the introduction:

- Logical criteria:
 - Spaces need to be defined such that they are unambiguously different
 - Distinction between spaces should be categorical, not gradual
 - Distinct spaces should involve distinct *goals*, and distinct *entities* being searched
 - What is “allowed” is passing of information from one space to others
- Empirical criteria:
 - Should take place at different *times*
 - Should involve different search *heuristics*
 - Behaviour should be influenced by different factors in each space
 - There has to be actual activity in every space
- Implementational criteria:
 - The spaces should be able to be represented in a computational model which can perform the task
 - Search through a specific space should be distinct from information passing between spaces

Since our description will not be sufficiently precise to allow the formulation of a computational model, we will put aside the implementational criteria for the time being. Judging from the logical requirement that spaces should differ categorical instead of gradual – as which differing levels of detail certainly would have to be considered – we have to assume that thinking with regard to *sheets* is bound to happen in one problem space, irrespective of possible layers of abstraction and detail.

Similarly all reasoning with respect to planning and drawing the *sketches* would have to be considered taking place in one space, whereas the possible dissociations between sheet and sketch knowledge we saw in the examples above certainly provide evidence that at least in some cases, subjects are thinking about sketches and sheets independently (with the possibility of information passing allowing them to coordinate, such as S6 did in several successive steps).

Since our study explicitly instructs subjects to keep those two activities apart, the empirical criterion demanding activity to take place at different times immediately suggests that we need different problem spaces for the subjects thinking about the paper folding task(s) and for them observing their thinking and fashioning their reports. Since introspection and reporting introspections verbally could be considered two separate activities, we tentatively use the plural, when considering these activities to take place in *meta-cognitive spaces*.

Finally, similar to the rule-induction tasks described by Simon & Lea (1974), we require spaces, which can take states of the sheet and sketch spaces respectively, and find regularities which could possibly connect them – *rule spaces*.

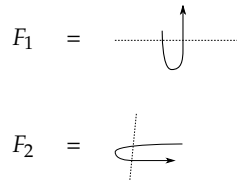
While there might be more problem spaces we could think of, we should first try and see how far we get with the ones we have up to now. Therefore, in the following sections we will discuss the possible formulation of the following spaces: A sheet space, a sketch space, two rule spaces, and meta-cognitive spaces. For each of these spaces we ought to discuss their *origin*, what their *states* are representing, what *operations* are being applied to these states, and which *goals* might be pursued in that space.

3.4.1 Sheet space

There is broad variety in the depth, form and abundance of thinking about the sheet, but given our task explicitly asks subjects to imagine it, this is bound to take place in some form or other. In Section 3.3.4, we described the associations subjects used to start thinking about our task, before the sketching sub-task was first introduced. We noted that these associations were with familiar experience, involving motor imagery of folding, coarse visual and spatial imagery of analogously folded objects (books, newspapers etc.), and metaphoric images embodying the folding process in images like the shoulders & arms one in Figure 29.

Representing sheet states

The first idea of how to characterise the folds could hence start out in motor imagery:



This indicates that the origin of a sheet space might lie primarily lie in the folding *operation*, with its state representations being recruited from familiar objects to which we can apply this operation. So the characterising property of the initial state representation is that the folding operation has been applied to it, with F_1 being characterised as an object which has been folded once, F_2 twice, etc.

$$\begin{aligned} F_1 &= \text{FOLD}(\text{obj}) \\ F_2 &= \text{FOLD}(\text{FOLD}(\text{obj})) \end{aligned}$$

We noted that the context of the associated folding experiences, specified the sheet orientation and maybe folding direction, so the FOLD operator must be able to amend this information to our state representations, e.g.:

$$\begin{aligned} F_1 &= \text{FOLD}(\text{dir}(\text{horiz}, \text{twrd}, \text{up}), \text{obj}(\text{align}(\text{horiz}))) \\ F_2 &= \text{FOLD}(\text{dir}(\text{vert}, \text{twrd}, \text{right}), \text{FOLD}(\text{dir}(\text{horiz}, \text{twrd}, \text{up}), \\ &\quad \text{obj}(\text{align}(\text{horiz})))) \end{aligned}$$

or with the application of a ROTATE operator between the the foldings:

$$\begin{aligned} F_1 &= \text{FOLD}(\text{dir}(\text{horiz}, \text{twrd}, \text{up}), \text{obj}(\text{align}(\text{horiz}))) \\ F_2 &= \text{FOLD}(\text{dir}(\text{horiz}, \text{twrd}, \text{up}), \text{ROTATE}(\text{dir}(\text{cw}), \text{FOLD}(\text{dir}(\text{horiz}, \text{twrd}, \text{up}), \\ &\quad \text{obj}(\text{align}(\text{horiz})))) \end{aligned}$$

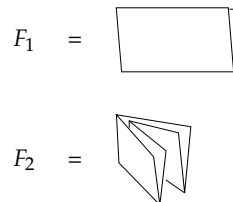
A problem with this style of successively including, *explicit symbolic information* in our fold state representations is that – certainly at the beginning of the task, but in part even in the later sessions – subjects do not have all this knowledge available⁵⁵, but it is *implicit in the way they are (mentally) acting*.

Making this distinction is very important, when trying to define *representational-level* problem spaces. On a computational level, we could employ all information we wanted to describe the subject's thinking and acting, on the representational level, we try to describe *in terms of the subject's knowledge*. For instance, computationally speaking, we could have described the fold state S6 was thinking about when he was planning to draw the sketch depicted in Figure 40 in terms of the BNF grammar for fold states defined in Section 3.2.6 as:

O:top P:right left-handed in-fold F5.

But S6 had done his thinking and drawing without any knowledge of chirality or the distinction of in- and out-folding, and in fact at several points in time, he was confused about whether his NV_5 was positioned on the left or the right of the sheet. He was “just” following the consequences of the folding procedure he had chosen to use in the beginning, and the sketch composition knowledge he had derived from there.

Generally, while we can *infer* from the way subjects speak about and sketch the sheet that it must be oriented in a certain way and that it must be folded in a specific direction, subjects often cannot immediately report this information when asked, but have to *return to imagery* to explicate it. This indicates that this knowledge must be encoded in a form which is not directly available for symbolic reasoning, such as the *depictive component* of mental imagery described by Kosslyn et al. (2006):



⁵⁵ Note that once this knowledge *is* available to the subjects, they also tend to make use of it, e.g. to make simple inferences about opposite sides of the sheet.

What seems to be more readily available than a fully explicit description of a fold's properties are descriptions in terms of *metaphorical images*⁵⁶:

$$\begin{aligned} F_1 &= \text{FOLD}(\text{dir}(\text{like-closing-a-book}), \text{obj}(\text{align}(\text{like-an-open-book}))) \\ F_2 &= \text{FOLD}(\text{dir}(\text{like-closing-a-book}), \text{ROTATE}(\text{dir}(\text{ccw}), \text{FOLD}(\text{dir}(\text{like-closing-a-book}), \text{obj}(\text{align}(\text{like-an-open-book})))))) \end{aligned}$$

Such metaphorical images might either be used as they are, as a sort of mnemonic short-hand for an explicit description, or as starting point in order to derive this explicit information.

So, regarding the representation of sheet states, there appear to be several kinds of representing the fold, all of which contribute some information to the subject's knowledge of the sheet. At the very least, these include motor imagery, spatial imagery, visual imagery, metaphorical descriptions and symbolic descriptions. The examples shown above only described very simple fold states concerning F_1 and F_2 as wholes, in order to describe a subject's more advanced thinking about the sheet, we would need to introduce ways to include reasoning about sides, edges, open and closed creases, etc. pp.

Manipulating sheet states

We already noticed how intimately states and operators are related in the case of our cross-folding task, so in discussing state representations, we already introduced the FOLD and ROTATE operators. Further actions which were applied to the sheet by pilot subjects include unfolding, changing perspective, focussing on single sides and tracing edges mentally with their fingers.

Introducing an UNFOLD and CHANGE-PERSPECTIVE operator seems not more complicated than FOLD and ROTATE (and had we defined a representational framework for fold sides, FOCUS-ON-SIDE could also be defined in a relatively simple manner) but it is really unclear what kind of *state transformation* could happen, when we *trace edges*⁵⁷.

But similar to how we needed to discuss the actions we want to apply to our states in the discussion above, we cannot talk about operators without discussing the kinds of representations they are supposed to manipulate. If our fold states were merely defined symbolically, defining our operators would be relatively easy. If a full symbolic fold representation would be a six-tuple containing a description of all 6 sides of the fold, we could formally define operators which would manipulate these tuples such that they are folded, rotated, and so on (as far as they employ knowledge available to the subjects, these definitions could be largely based on the computational descriptions in Section 3.2).

But defining operators becomes much more difficult, if we consider all the different forms of representation discussed above. Folding a symbolic description, a metaphoric description or a visual depiction are three very dif-

⁵⁶ Note this example's orientation and folding directions do not correspond to the examples above.

⁵⁷ Following our interpretation of the pilot cases finger tracing might not so much be an operation concerning the manipulation of sheet states, but an operation which allows thinking about the sheet and drawing edges at the same times. Thus, in terms of problem spaces, we could suppose that the TRACE operator plays a larger role in setting-up the sketch space.

ferent operations. And even if Schunn & Klahr's criteria assure us that different forms of representation should be placed in the same problem space, as long as they represent the same object and are used in aid of the same goals, according to Klahr & Dunbar (1988, p.38f) *implementationally* we might actually be forced to place these different forms of representation all in their own problem spaces and supply them with complicated interaction mechanisms.

While it can be hard to work out from empirical data, we should also check whether some operator might actually only be attached to a certain, preferred form of representation (e.g. changing perspective seems to be something that mostly happens in visual imagery, whereas spatial and symbolic reasoning usually deal with canonical orientations and perspectives). This is not only an important step to save us from unnecessarily defining operators which are not actually used, but is a crucial point regarding the question *why it is advantageous to think employing multiple different forms of representation*. Apart from occurring with one form of representation and not with another, it might also be the case that some operations require minor mental effort in one form and are very difficult in others⁵⁸.

Goals

In terms of sheet spaces, the goal induced by our task of imagining the fold and acquainting oneself with the resulting object, is a bit unusual in the context of problem solving, since our goal is one of *operator application and instantiation*. What subjects need to do, is to apply the FOLD operator a given number of times, and find out what exactly the *consequences* of this application were. In contrast, most problem solving tasks usually concern the subject having to decide between several possible operators, in order to attain a defined goal state to which they have to find a solution path hitherto unknown to them.

However, applying the FOLD operator still poses a problem to the subjects, insofar as they do not know precisely what its consequences are, and what good ways of representing the results would be. In other words, we might consider our task to pose a problem in a problem space hierarchically superordinate to the sheet space – with the aim being to develop the sheet space such that the FOLD operator can be applied successfully.

Regarding implementations

Wintermute (2009) and Laird (2012, chap.10) describe a system called SVS ("Spatial/Visual System"), which was added as an extension to the Soar architecture in the late 2000s. This system integrates visual and spatial imagery into Soar's architecture, and allows Soar to reason about problems *in a multitude of representations*, namely: amodal symbolic, quantitative spatial and visual depictive. These can all be used to represent different aspects of a problem, and crucially all of those can interact with Soar's working memory and central deliberative processes, i.e. they all can contribute to problem solving and learning. To our knowledge this is one of the few systems which

⁵⁸ For instance, there are many discussions regarding the ease, but also error-proneness of employing diagrammatic forms of representation in solving mathematical and physical problems (Larkin & Simon, 1987; Koedinger & Anderson, 1990; Raab, 2015b).

tries not to model visual imagery or symbolic reasoning in separation, but tries to elucidate how they could work together⁵⁹.

Our descriptions above are far too vague to approach an implementation of a sheet space in Soar/SVS for the time being. However, if anything like this should be attempted in the future, it would require a system which allows states of affairs to be represented in multiple forms, and allowed these forms to interact and contribute to a common problem-solving process (as does Soar/SVS).

3.4.2 *Sketch space*

Analogously to the sheet space, development of the sketch space starts out in subjects through one of our task instructions – in this case with the first sketch task (T1.3). At this point subjects will already have done some imagery and reasoning regarding the folded sheet, however our observation from the pilot cases were, that thinking about the sheet in greater detail only really began with the presentation of this task. Also similar to the sheet space, which is originally constituted by the action of folding, the sketch space is build based on a physical activity – drawing.

As opposed to the initial situation with the sheet space, however, our pilot subjects did not report many associations to memory episodes of drawing, motor imagery or metaphorical images concerning drawing. We assume that this is (a) due to the sketching task giving very precise instructions, and (b) due to the first cross-folding task (T1.2) already constraining the subjects' expectations of what possible follow-up tasks there might be. Hence, they have lesser need to employ associations to remote knowledge to organise their thinking, because they are already focussed on a subject matter.

Representing sketch states

Generally, we might assume that sketch space starts out with states representing arbitrary, 2-dimensional arrangement of *pen strokes* (even if some of these strokes were meant to depict three dimension through the use of perspective). This includes, but is not necessarily constrained to *visual-depictive* representations. We could equally well conceive of them as being based on *motor imagery* of moving a pen with our hand (although, as mentioned above, there is little evidence for that in our pilot data). Additionally, in our pilots we witnessed a lot of complex reasoning about the sketches composition (see Section 3.3.7), which means we have to introduce sketch representations (or rather: representations of the sketches' composition) which allow *symbolic reasoning* and the discovery of *metaphoric images*.

So, since it involves all forms of representation we also assumed for the sheet space, we might conceive of the sketch space as a 2-dimensional version of sheet space. Yet, there are major constraints to what is being represented in sketch space, and allowing *arbitrary* 2D sketches would open up space way larger than what we need to describe our empirical observations.

T1.3 instructs subjects to fashion “a schematic sketch of F_2 seen from left, right, top, bottom, front and back (6 2D-sketches in total)” and to mark

⁵⁹ Related ideas are discussed in Davies et al. (2011) and Croft & Thagard (2002).

which view is which. This includes very specific information which should inform how sketch space will be set up:

- The sketches ought to be schematic (excessive realism and precise measurements do not matter)
- The sketches ought to be 2-dimensional
- There have to be 6 sketches in total (left, right, top, bottom, front and back)

If the task has not been misunderstood, we should expect representations of sketch states to be 2-dimensional and *qualitative*, i.e. disregarding metric measures, and instead only relying on relative magnitudes. Indeed, all our pilot subjects drew 2-dimensional sketches, and their coarse, but bounded use of stroke lengths, angles between edges etc. suggest that they did neither put them arbitrarily nor quantitatively reason about them.

Sketch representations can come with the knowledge of which of the fold's sides they depict, but this does not have to be the case, since sometimes, subjects had to return to lengthy episodes in imagery to find out. So, we should allow annotating the *position on the sheet* to a sketch representation with the option of keeping this property undetermined.

Additionally, we have to consider *differences between the representations and the actual sketches*: Even though subjects are instructed to mentally plan fashioning the sketches in every detail, they regularly encounter unexpected difficulties when drawing, mostly regarding the *spacing* of strokes or sub-figures. This might be seen as indication that subjects less frequently consider visual-depictive representations of sketches drawn mentally (which should allow them to anticipate the difficulties of drawing), but that more often they compose *descriptions* of the sketches they plan to draw.

The possibility of composing sketches of impossible physical situation we have seen in some of our pilot cases above indicate that the sketch space does not necessarily have to be heavily constrained through sheet knowledge. Interestingly however, this is a phenomenon which usually occurred in later session, when more complex folds had to be drawn. With the exception of two subjects (S2 and S6) misunderstanding the task instructions when drawing their first sketches, every subject fashioned correct sketches of F_2 in their first attempt at T1.3 (and S2 and S6 did so after their misunderstandings had been clarified). This could be due to F_2 still being simple enough to be imagined in sheet space, rotated to the according side (or the side being otherwise focussed) and traced with mental pen strokes side-for-side⁶⁰ – whereas for the sketches of $F_{\geq 3}$ subjects need to structure and decompose the sides.

So, over the course of the sessions, subjects discover the primitive forms (I, V, DV, NV) and several means of composing them: Putting-side-by-side/*stacking*⁶¹, *nesting*, and *creasing*⁶².

With the availability of this knowledge, we could suppose that sketch space undergoes major *representational changes* – from sketch drawing to sketch composition. Instead of exploring the sheet in sheet space and depicting it

⁶⁰ A highly speculative interpretation.

⁶¹ Depending on the orientation of the sheet. For reasons of terminological simplicity, we will remain with the term “stacking”.

⁶² In order to avoid confusion with folding the sheet, we will refer to folding of a 2-dimensional figure as “creasing”.

bit by bit with mental pen strokes, subjects compose the primitive forms in order to re-create the sheet's sides, and bringing forth the array of impossible sheet situations depicted in some of the sketches we have seen (e.g. Figure 42, 43). This sketch composition approach has then to be successively constrained by sheet knowledge, and could possibly ultimately end up with a variant of the correct, general procedures for sketch composition described computationally in Sections 3.2.9–3.2.11⁶³.

Manipulating sketch states

As with the sheet space, the operators we have crucially depend on the kinds of representations we are employing. We already noticed that there are at least two broadly different “modes” in the sketch space – sketch drawing and sketch composition (with both having several possible underlying forms of representations).

Regardless of which mode we are using, we rely on returning to sheet space whenever we lack knowledge about how a sketch should possibly look. This could be described by a RETRIEVE-SHEET-KNOWLEDGE operator, which gets called when we do not know how to draw (or how to compose) a sketch, and fills in this lack of knowledge, for instance by calling the TRACE operator in sheet space.

Also, there is the possibility of *manipulating* sketch representations, e.g. with a ROTATE operator, yet there were only very few cases in our pilots, and most of the time subjects ignored properties such as orientation when being concerned with sketches of a single side of the sheet.

With sketch drawing, the main activity seems to be making pen strokes (regardless of whether they are represented visually depictive, in motor imagery, or purely descriptive), thus we might posit the existence of a STROKE operator, which generates these representations.

We already mentioned the three main activities of sketch composition above, and they also constitute the operators for this mode: STACK, NEST, and CREASE. The STACK and NEST operators are relatively simple to implement mentally, since they are relatively independent from the from complexity of the sketches they ought to stack and nest – simply take two old sketches and put them on each other, or one within the other. But in case of the CREASE operator, carrying out that action can get very difficult with increasingly complex sketches. For instance, in his second session, S6 explained his approach at creasing sketches as “halving and creasing them”, providing this illustration:



Figure 46: S6's illustration of his “halve & crease” approach.

Faced with ever more complex sketches in later sessions, S6 developed the more complex creasing procedures which we described as an example of

⁶³ Interestingly, (Ohlsson, 2008; Ohlsson, 2012a; Ohlsson, 2012b, chap.7) describe a very similar situation – the successive specialisation of overly general procedural rules by observing mismatches between declarative constraints and action outcomes in the context of cognitive skill acquisition.

systematic sketch generation procedures in Section 3.3.7 (and which might ultimately lead to a general procedure of sketch composition such as the ones described in Sections 3.2.9–3.2.11).

Especially when faced with having to crease a sketch or sketch component which cannot be understood as easily as the example in Figure 46 (which could e.g. be understood without imaginary creasing as “mapping an I to a V”), sub-figures sometimes have to be *decomposed* again, i.e. trying to see a V once more as an arrangement of two strokes. This might be achieved by a DECOMPOSE operator. Note that re-interpretation is one the main pieces of evidence brought forward by the contenders of visual-depictive theories of mental imagery, since such interpretative changes are hard to imagine without appealing to visuo-spatial properties of an image (Finke et al., 1989). Hence, the activity of the DECOMPOSE operator might not manipulate any sketch representation directly, but mediate this change by switching between different kinds of representation.

Goals

The main goal in sketch space is very close to our formulation of the drawing task (T1.3, 2.3 etc.), namely to find those sketch states which correctly represent the sides of the fold state currently asked for. This might be achieved by finding the right sequence of applying the STROKE, NEST, STACK, and CREASE operators, and by calling the RETRIEVE-SHEET-KNOWLEDGE operator when there is too little knowledge to decide how to proceed with composing the current sketch.

The successful retrieval of sheet knowledge should allow to subsequently integrate this knowledge as in the form of *constraints* on the application of the other sketch space operators. Thus, this interaction might not solely take place between the sheet and sketch spaces, but also involve the composition of possible rules in both *rule spaces*.

With more complex folds, an important goal which must be added to sketch space, is that the composition of the sketch looked after must not only be correct, but given in a form such that can be drawn easily and without making too many errors⁶⁴.

Regarding implementation

While there is a lot of work on the computational modelling of sketch *understanding* (e.g. Forbus et al., 2011; Krumnack et al., 2013), *producing* sketches has received far less attention⁶⁵.

A starting point might be provided by the CogSketch system described in Forbus et al. (2011), which is a representational framework for sketches combining visual, spatial, and conceptual knowledge. Its sketch representations are built from “glyphs”, consisting of visual polylines and symbolic tokens,

⁶⁴ If we want, we might also claim that this sub-goal or added goal constraint was always there, but only becomes a relevant factor with later folds.

⁶⁵ A rare discussion of the topic is provided by Goel (1992, 1995) which considers sketch generation at length, yet they conclude that its underlying processes are fundamentally incompatible with “rigid” computational models such as proposed by problem space theory. For a developmental perspective on the cognitive complexity of drawing and sketching, cf. also Machón (2013).

and can be structured hierarchically by defining additional symbolic layers of abstraction, qualitative relations between sub-sketches etc. Since CogSketch is primarily meant to be an interactive tool for research on sketch understanding, and educational tutoring it lacks facilities for autonomous drawings and only works on user-defined glyphs.

3.4.3 Rule spaces

We already mentioned the idea that the exploration of possible ways to structure and understand (i.e. represent) the folding domain can be seen to some extent to be main goal of our task. Following this idea, we would have to describe major parts of this activity in terms of rule space activity (which in turn induces changes in the sheet and sketch spaces). However, in trying to define the rule spaces, we are faced with several difficulties:

First, since they are supposed to take sheet and sketch space activity as their inputs and change them according to the regularities they were able to identify in this activity, they are hierarchically super-ordinate to these spaces. Hence, they “inherit” all the ambiguity we were already facing when trying to define those. So, however we might end up defining the sheet and sketch spaces, our definition of their rule spaces has to accommodate these definitions. In other words, we do not only have to deal with the uncertainties of defining rule space activity per se, but that these definitions are being complicated further by having to accommodate all different possibilities of defining the sheet and sketch spaces.

Second, in the tasks modelled with existing multi-space layouts we mentioned in the introduction, such as Simon & Lea (1974) and Klahr & Dunbar (1988) *finding the rules* governing a presented task domain was their subjects’ *explicit* task, whereas in our case, even though we said above that finding regularities was the main goal of our task, this is something we have in mind as investigators – actually it is more of a hidden agenda and having to find regularities is a side effect of having to deal with the task’s complexity.

This means that our subjects rarely explicitly reasoned about what regularities and invariants could possibly be found in the paper folding domain, rather they *encountered* these regularities while struggling through the task of repeatedly imagine the folds and plan their sketches. Accordingly, we had to *infer* most cases of rule space activity from observing the effects of having encountered a regularity in the subjects’ reports, but lacking a greater number of direct reports about such activity there is not much we can say about how exactly it was taking place. This might be due to our setup failing to instruct subjects to report this activity or due to larger parts of this activity being difficult to report, e.g. because it takes place as *implicit learning*, bringing forth *tacit knowledge* (Reber, 1993, cf.).

Nevertheless, our observations in Section 3.3.6 allow us to derive some characteristics of rule space activity. We noted that our subjects were identifying regularities in the form of *rule-like regularities* as well as *metaphorical images*. These regularities were *between-folds* or *within-folds* and described *predicative* as well as *functional relations*, with functional relations being more useful for our task domain. Metaphorical or analogical images could be based on su-

perficial similarities (e.g. S5's memory of her fax machine) or on structural similarities (e.g. S1's Dracula)⁶⁶.

Presenting their "representation space", Schunn & Klahr (1995, p.4) describe three mechanisms of representational change which they inferred from their data: *Notice invariants*, *analogy* and *brute-force search*. *Notice invariants* allows subjects to notice repetitions and regularities when observing the outcomes of their sub-ordinate problem space's activities – for instance to notice changing vs. unchanging parts of some outcome. Subsequently new representations can be chosen which emphasise this regularity. A fitting example from our pilot data would for instance be subjects' leaving out the two rectangular sides from their sheet representations (and later on the I and the V).

The *analogy* mechanism Schunn & Klahr propose is explained as producing representations by analogy to previously understood phenomena, applying features of the analogical source's representation to the current ("target") representation, and hence leading to a re-categorisation. The circumstances under which such a process might be triggered are the occurrence of salient expectation-violating events (ibid.). This is in accordance with Chan et al. (2012) who observed that analogy is used particularly often when dealing with uncertainty. S3's metaphorical image of the mountain massif may be a good example for this process: It first occurred when he described a salient, but ephemeral visual impression (the DV's "massivity"), but later on he could use this image to re-structure his sheet representation by elaborating on the edges building the "crest" joining the two Vs of the DV in the background, and the "valley" between them.

The last mechanism proposed by Schunn & Klahr is *brute-force search*, which they describe as the subjects just trying out any form of representation they have available for the object in question. As opposed to the first two, we could not find pilot data which could be well described by using this mechanism. This might be due to the ill-defined nature of our task: Since our subjects do have to come up with their own way of representing the sheet and the sketches – without a normative form of representation being suggested to them – switching between representations in a trial-and-error manner might not look like a promising strategy.

3.4.4 *Meta-cognitive spaces*

The meta-cognitive spaces face the same problems as the rule spaces: They are hierarchically super-ordinate to the problem-level spaces and their activity can be difficult to infer from the data. In our general template-layout, we assume two meta-cognitive spaces: A general *meta-cognitive space* and a *report space*. This distinction is made on the observation that (a) making introspections and reporting about them are two distinct activities (we can introspect and not report about it)⁶⁷, and (b) that our reports contain other observations than introspections, e.g. problem-level comments, hypothesised

⁶⁶ This distinction is independent of whether the image turned out to be false or unhelpful for subsequent problem solving. In this regard, Wiener (2006) distinguishes *potent* and *impotent* analogies.

⁶⁷ Note however, that through the very nature of our data, evidence for introspective activity by our subjects is *necessarily* mediated by them *reporting* about it, and even though we distinguish a meta-cognitive and a report space logically, we would require hitherto unavailable methods of data collection to make this distinction empirically.

rules, general commentary, etc. (see also our tentative coding scheme for report types in Section 2.11.8).

3.4.5 *General meta-cognitive space*

Since introspection and meta-cognition (or cognitive monitoring) are additional thought processes in themselves and go beyond the mere verbalisation of sub-vocal language-like thoughts (Ericsson & Simon, 1998), we place these processes in their own problem space. Yet, we do not know much about these processes, except that they involve directing attention to one's own cognitive activity and that by allowing to notice patterns in this activity, they can be responsible for changes in task comprehension, strategies and representations, and subsequently for increased task and transfer performance (Cox, 2005; Dominowski, 1998; Fox et al., 2011).

Furthermore, this process is fallible and can involve the confabulation of self-explanations *a posteriori* (Nisbett & Wilson, 1977). In order for introspection or meta-cognition to be able to attend to a cognitive process, this process has to be available to *consciousness* (i.e. we cannot attend to unconscious processes). Acknowledging the confusion and general lack of theoretical foundation for what constitutes consciousness and what decides which processes are conscious and which are not, we will not attempt a further description of the general meta-cognitive space in this thesis⁶⁸.

3.4.6 *Report space*

Fashioning the written reports is a large part of our study. Hence it should be expected that our subjects spend a long time with the composition and writing of these reports. As we mentioned above, reports can be meta-cognitive and introspective, but they do not have to, and might equally well describe other spaces' states directly. Hence, we have to assume that a report space, can take input from any other space in our template-layout.

Apart from the report space's main activity – composing linguistic descriptions – it can have an important influence on the other spaces' activity: We already noted examples of metaphoric images and similar constructs that were used by the subjects with *communicative intent*, i.e. in order to express a thought anticipating that the reader (the investigators) might have difficulties understanding the report otherwise. Judging from his data, S3's mountain massif was very likely an image used communicatively, and – as we already described above – used productively by S3 at later stages. So, we might assume that the report space includes an operator which requires more detail about another space's state (or alternatively: a description of that state in a form of representation which is easier to describe linguistically⁶⁹).

⁶⁸ While there is a multitude of theories about what consciousness might be in general, none of them is detailed enough to be applicable in a concrete model such as that of our task.

⁶⁹ This would be a case of verbal overshadowing – so if the subject still remembers our task instruction, they should in this case also be able to draw a sketch as alternative means of expressions.

3.4.7 The whole layout

This concludes our introduction of tentative representational-level problem spaces for describing a subject working on our task. We distinguished between a sheet space, a sketch space, two rule spaces (one for sheet space, and one for sketch space), a general meta-cognitive space and a report space.

Each of these spaces allows *multiple forms of representations* and their operators either work adaptively with each of these forms or are only available for some of them, invoking *changes* of representation when they have to be applied. The spaces are arranged in a *hierarchy*, such that super-ordinate spaces can induce representational changes in sub-ordinate ones. General exchanges of information (possibly inducing changes in strategies, goals and other more malleable components of a problem space) are possible in both directions.

General *interaction* is possible between almost all spaces, although some interactions are far more likely than others⁷⁰. The only interaction which we did not consider explicitly was between sheet space and the rule space (sketches), as well as the other way around. Even if we could think of a possible situation in which such an interaction might take place, it makes more sense to assume that it would be mediated by sheet space or sketch space, respectively.

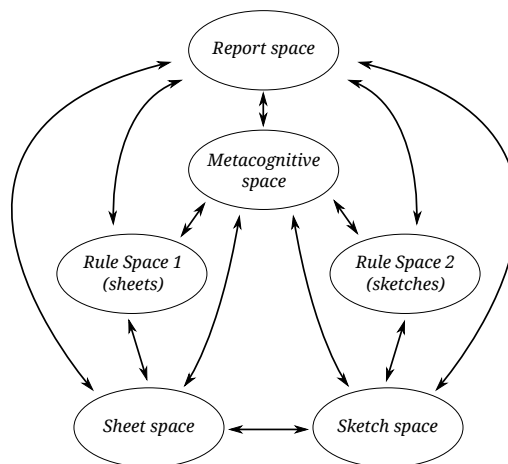


Figure 47: A 6-space “template” layout for describing our task. Every space except the two rule spaces, and a rule space with the opposite problem-level space may interact with each other.

3.4.8 Layout variants

Note that even though at some points we mentioned possible differences between the individual subjects’ thinking, up to here we talked about these problem spaces (or “templates”), as if they fitted every subject and every point in time equally. To get an impression of the variability between individual subjects as well as between different points in time for the same

⁷⁰ E.g. sketch and sheet space can interact very intensely, whereas every space is connected to the report space because it is possible that such “unreflected” reports will be made, but we still expect the majority of reports to derive from the meta-cognitive space feeding into report space.

subject, we describe a few modifications of the general layout presented above as examples.

The examples given below only concern the relation of sheet and sketch space. It is likely that there would be similar amounts of variability with regard to the rule and meta-cognitive spaces, but due to their more speculative nature, we do not really have a basis for describing any such variability in our pilot data.

Sheet space develops into sketch space

We noted in our formal descriptions of the cross-folding domain (Section 3.2) that we can fully determine a fold state by describing its sides, and even by only describing its NV together with its orientation and position on the sheet. We can imagine a subject making a similar development (although we would assume, they would need more than 4 sessions for that). Once the operators of a sketch space have fully integrated the constraints given by the folding operation of sheet space, operations in both spaces are coherent, and it is possible that subjects will cease considering both ways of thinking about the sheet.

Since even with advanced knowledge and the employment of smart metaphors it becomes increasingly difficult to imagine the sheet visuo-spatially (at later points even imagining one side could become too complicated), it is possible that a subject will cease referring to imagery at all at some point (and given that they gathered all the principle domain knowledge from imagery, they might not deem it necessary to keep trying). This development might be described as the subject having developed their sketch space to fully take over the role of sheet space – with talking about sketches being just their way of talking about folds.

Dominant sheet space

Irrespective of the space layout that may develop at some future point in time, we encountered subjects for whom operation in their sheet space remained the primary task for the whole duration of the pilot study (we already alluded to this possibility with Figure 31). Fashioning the sketches was regarded as a secondary task and reasoning about the sketches took place in a sketch space which was almost completely dissociated from sheet space (leading to a sharp increase in sketch errors after the sheet became too complex to use naïve tracing as strategy).

Whereas other subjects also used sketch representations to reason about the sheet (as far as they had integrated the respective sheet knowledge into their sketch spaces), these subjects remained with working in sheet space and imagery for as long as they could, and only deferred to thinking about sketches, when they had to draw them (i.e. the exchange of information between sheet and sketch spaces was only unidirectional for them).

Dominant sketch space

Conversely, there were subjects who concentrated on the sketch task pretty quickly, and generally avoided reasoning about the sheet. This might cor-

respond with a general tendency to avoid complex visual imagery, which plays a large role in sheet space before subjects have derived conceptual knowledge about the sheet⁷¹.

Since the whole task is conceptualised as being concerned with fashioning the sketches, these subjects tried very early on to rely purely on sketch knowledge, even if they had not yet integrated all the sheet knowledge necessary (leading to an increase in sketch errors at points in time where a solution employing sheet imagery would have been still possible). Since they still had to integrate sheet knowledge at some point, it would be an exaggeration to claim that information exchange between the two spaces was unidirectional (with the opposite direction as in the example above), this integration took place significantly later and to smaller amounts than for other subjects.

3.4.9 Concerning representational-level problem spaces

This concludes our discussion of representational-level problem spaces with regard to our task. We noted several *difficulties with defining these spaces*, mostly related to (a) the lack of sufficiently informative empirical data, (b) a lack of theoretical foundations regarding meta-cognition, introspection and consciousness, and (c) implementational difficulties with regard to multiple and possibly indeterminate representational forms.

We included several examples for possible differences between subjects, as well as within one subject over time. While we can respond to differences between subjects with adapting our generic template-layout of problem spaces to the individual thinking of the specific subject in question, changes *within* one subject would have to be explainable from within this set of problem spaces, i.e. their rule and meta-cognitive spaces would have to induce the changes that e.g. lead to the merging of their sheet and sketch spaces.

Ultimately, we have to note that a definition of representational-level problem spaces based on the *commonalities* between our subjects can still explain some of the variety between subjects, they cannot be formulated such that they explain an individual's thinking in sufficient detail: Either this generic model would have to be formulated so broadly that it could accommodate each individual subject's thinking – and thereby assume problem spaces of which each individual would only use and explore minute parts of⁷². Or we would have to abstract away large parts of each subject's idiosyncrasies, and give up the aspiration to explain these aspects of their thinking with problem space-models at all.

3.5 CONCLUSION

In this chapter, we presented the current state of the *task analysis* for the cross-folding task from our study. It was based on a survey of *related research* in problem solving (already included in Chapter 1), mental imagery, paper

⁷¹ And thus, reasoning about sketches at least compartmentalises this imagery task regarding one complex 3D-object into 6 smaller 2D-objects, constructed from arrangements of edges or strokes.

⁷² Begging the question why they would possess this knowledge and skill in the first place, if they did not use it or even consider using it.

folding, and cognitive skill acquisition, a presentation of several *computational-level descriptions* of the cross-folding domain, an overview of *observations from our pilot studies*, and the formulation of tentative *representational-level problem spaces* based on these observations.

The aims we formulated for this presentation at the beginning of the chapter were (a) to show that our task fulfilled the task criteria put forth in Chapter 2, (b) to provide a framework for the interpretation of future data, and (c) to form the background for a discussion of representational dynamics and problem space theory.

We achieved the first aim with discussions of all 5 task properties at appropriate places throughout the chapter. The second aim concerns future work and thus we cannot claim to have achieved it. However, our use of the formal descriptions of the task domain in presenting the pilot observations, as well as our use of these observations for defining the problem spaces should have provided an indication of the utility of these descriptions. We thus assume that the work presented in this chapter constitutes a considerable resource for the interpretation of future data. Finally, the discussion alluded to in the last aim takes place in the following, concluding chapter.

DISCUSSION

"I still feel as though I'm baring my soul to my mind here."

Utterance B171 of Subject S3 (Spring 1960) — Newell 1967, p.143

In the introduction, we formulated the question "Where do problem spaces come from?", and further specified it as "How are they constructed? Why are their constituents chosen as they are, why and when do they change, and what mechanism mediates this set-up and change?". We diagnosed the current lack of knowledge to stem from a concentration of research on WDPs, problems which are relatively easy to comprehend and which in turn involve little necessity for learning and representational dynamics.

We proposed a task and study set-up which we assume could help to shed light on these phenomena and aid subsequent theory-formation. The two main deliverables of this thesis were to present the *design of this study*, and a first *task analysis*, based on (a) sketches of computational-level descriptions of our task-domain, (b) a selection of observations about representational-level structures in our interpretations of our pilot data, and (c) an attempt at synthesising these observations into specifications of representational-level problem spaces.

So, having done all this, can we answer our initial questions? Given the wide scope of these questions, it should be no surprise that the work presented here cannot constitute a final answer to these questions, but only provides a small contribution to a hopefully converging body of knowledge. Therefore, in the following we will discuss several pointers for *theory formation* as well as *future empirical work* which we derived from this work.

4.1 POINTERS FOR THEORY FORMATION

In the introduction, we noted that Simon (1973) proposed that even for IDPs, people always work in well-defined problem spaces, with these spaces changing through *additional processes* of recognising the need for new operators or state representations and retrieving them from memory. Multiple problem space models can be seen as further developments of this general idea. However, even though we presented first steps towards a multi-space model of our own, there seem to be several principled questions these models leave unanswered:

Explaining representational dynamics

Hierarchical multi-space models try to explain changes in problem spaces by assuming super-ordinate spaces in which these changes are searched for and induced. Thus, the sub-ordinate space (usually the one concerned with the actual problem posed in a task) is a *changeable entity*, reflecting

the effects of learning, insight and the general growth of knowledge and expertise. The organisation of the super-ordinate spaces on the other hand is assumed to be unchanging and fixed: A rule space includes a fixed set of operators for hypothesising rules, leading to a model of an unchangeable learning process working on top of a changeable problem-solving process – with both of these processes being described in terms of a problem space.

Yet, we definitely could observe differences in the type of regularities that people tried to look for in our task: For instance, why does someone look for visual metaphors in one situation, but for syntactical rules in another? Burns & Vollmeyer (2000) proposed a rule space which can be changed by a *model space* (in which the task's goal is supposed to be modelled) – so changes in rule space should be explainable by observing changes in a subject's task comprehension. But while this could possibly explain cases where two subjects' different comprehension of our task lead to them looking for different kinds of regularities, it still cannot explain how *the same subject* starts looking for different kinds of regularities without changing their outlook on the task (which we could also observe)¹.

In our pilot studies, we could observe (a) that subjects used different forms of representation, (b) that they did not employ these forms exclusively, but in an interactive manner, and (c) that the initial selection of possible forms of representation were partly dictated by associating familiar knowledge, situations, and memory episodes to their reading of the task instructions.

Among others, this leads to the following questions which future research will have to answer:

- Which process decides what form of representation to employ from a moment to moment basis?
- How is information (partly) conserved when changing between different representations?

Is it really 'problem spaces all the way down?'

As opposed to the later multi-space models, in his original article on ill-defined problems Simon did not consider the additional processes which change the problem space to be organised as problem spaces themselves. Instead, he described them as (putatively unchanging) *parts of the cognitive architecture*, referring to general capabilities such as feature recognition and memory retrieval (Simon, 1973, p.191f). This leads us to two further questions:

- Should we always assume (yet another) super-ordinate problem space if we can observe change in a problem space?
- If we go along with Simon, should we reserve problem spaces for the description of changeable cognitive contents or also describe the activity of fundamental, cognitive processes in the same conceptual framework?

¹ While there might be a rough developmental trajectory of subjects starting out to deal with our task more in visual, spatial and motor imagery and ending up thinking more in terms of rule-like regularities and symbolic knowledge, there were no separated "phases" of thinking about the task either in this way or in that way. During any point in time, subjects might change between different forms of representation (with considerable, but never perfect transfer of knowledge between these forms).

The first question hints at difficulties in following all of Schunn & Klahr's criteria for introducing new problem spaces *at the same time*: Sometimes we can observe activity which when described formally appears logically detached from one another, implying we should describe them in two spaces. But empirically this activity seems to happen almost simultaneously (or we can only infer that it must have happened recently), indicating we should describe them in the same space. The implementational difficulties which start to appear when trying to describe how several forms of representation interact with each other in the same problem space, and are being manipulated by the "same" operators bring along even more difficulties (cf. Klahr & Dunbar, 1988, p.38f; Minsky, 1992).

This could mean that we either have to *loosen the criteria* and e.g. introduce problem spaces for which we do not have an empirical basis, or alternatively assume that *some changes appear to be caused by processes not organised in problem spaces themselves*. This latter option suggests itself because our observations make it hard to stick to the assumption of a *single, sequential locus of activity*. If we allow that there are multiple, changeable problem spaces, but require that moment-for-moment we are always situated in only one of these spaces, it gets hard to explain phenomena like implicit learning (Reber, 1993), the apparently unconscious retrieval of analogical sources and other associations (Hadamard, 1954; Hofstadter, 2001), the effects of implicit processes such as priming (Schunn & Dunbar, 1996), or the modification of memories through analogical reasoning (Vendetti et al., 2014).

This brings us to the second question formulated above, which we could have also phrased as "Which kind of processes do we want to model with problem spaces?". The Soar architecture proposed to describe *all* cognitive activity whatsoever in terms of problem spaces. On the other hand, in problem solving research, a common (if rarely expressed) assumption is that problem space activity is only supposed to model sequential (potentially verbalisable) activities in working memory², which in turn often are associated with conscious thought (cf. Baars & Franklin, 2003; Baddeley, 2007, chap.16).

But our observations, as well as further evidence of the unconscious underpinnings of association, analogical retrieval, memory modifications, and implicit learning (see above), strongly indicate that some of Simon's additional processes appear to work neither consciously nor sequentially. Therefore, we find ourselves faced with the choice of either assuming that non-problem space activity is involved in problem solving, or to give up the assumption of a sequential single locus of activity and the identification of problem space activity with conscious thought. Considering that the latter option would make problem spaces a far more hypothetical entity, and that we would need to explain why some problem spaces describe conscious thought while others do not, we find the first option to be the more parsimonious strategy for theory development.

Furthermore, by assuming that our thinking always takes place in a well-defined space, Simon's answer and the subsequent multi-space models developed in its spirit underplay occurrences of *uncertainty, ambiguity, confusion, mind-wandering* etc. If they were only distractions which maybe influenced problem solving *performance*, but had no further reaching influence on

² E.g. Newell (1979, p.4) claims that problem space activity models "human goal-directed symbolic activity" and requiring that activity must take place serially (ibid, p.7).

the course of our thoughts, we might explain them as mere imperfections of human cognition approximating an idealised problem space-architecture. But as it turns out, all of these phenomena *can* influence our task-related thinking and could potentially be productive for our solution process³ (cf. Section 3.3.1; Schreiber, 2015).

So, if our theory cannot account for these phenomena it has to be considered incomplete. And ultimately, this provides an indication that maybe we *cannot maintain the existence of perfectly separated spaces* in which problem solving takes place, since these spaces are influenced by “task-unrelated” perception, thinking, emotions and also bodily functions such as tiredness, hunger and many more (cf. e.g. Barth & Funke, 2010; Isen et al., 1987)⁴.

Possible directions which could be considered for theory development regarding the problems discussed above are (a) taking into account the motivational and affective modulation of cognition beyond the goals provided by the task instructions (Bach, 2012; Minsky, 2006; Simon, 1967), (b) the inclusion of implicit, associative processes in some form of dual-process theory (Sun et al., 2005; H  lie & Sun, 2010), and (c) an investigation of the phenomenological concept of *background* described in Agre (1988, chap. C6)⁵.

4.2 FUTURE WORK

Our discussion of future work is divided in two sections. First, we discuss concrete consequences we took from our pilot studies, and which should be used to further develop our study design presented in Chapter 2, and second we discuss open questions which could be used to guide future research.

4.2.1 Lessons from the pilot studies

There were many smaller lessons taken from conducting the last pilot studies⁶, but the main observation we made was that even with our extensive and well thought out training session the *quality and resolution of our data still varied widely*.

While for all pilot subjects, the quality and resolution of their reports clearly increased through the training session, there were large differences in the *type of information* reported (e.g. some described more phenomenologically, some more on the functional task-level), as well as in *time resolution* (ranging

³ In light of the over-reliance on distraction and coincidence by popular accounts of creative thinking, it is important to note that this productivity is possible, but of course not guaranteed.

⁴ Historically, this strong emphasis on a separate, specific reaction to being presented a task, and working on this task in a goal-directed manner was a reaction to the undirected constellation of responses put forward by early associationist psychology which had been struggling to explain goal-directed thinking. Selz (1927) provides an early account of the ideas which later developed into the concept of the problem space and means-ends analysis (cf. Mandler & Mandler, 1964, chap.6 for an English translation).

⁵ Agre discusses how situations are to large parts co-constituted by many elements which are not reflectively processed, but which shape a situation by suggesting some courses of action and averting others. Oswald Wiener’s notion of *orientation* captures a similar intuition, but without implying such a strict distinction between reflected foreground and pre-reflective background (cf. Wiener, 2008; Raab & Eder, 2015).

⁶ E.g. we noted that the word-morph problems used as parts of the warm-up tasks in Session 1 (T1.1) were surprisingly hard for some subjects, indicating that we might have overestimated native German speakers’ active knowledge of English three-letter words.

from one data entry every 5 or 10 minutes to several data entries per minute) and in *how much detail* was reported.

Given that our current training session already contains a lot of information for subjects, it would possibly not make sense to include even more instructional detail in it. Instead, it might be beneficial for subjects to simply have *more time to gather experience* with problem solving in imagery and reporting. One possibility would be to introduce *a second training session* after the first one, in which subjects would work on other (possibly already more complex) training tasks before starting with the first cross-folding session.

Since the problem of data quality and resolution is at its most severe right at the beginning of our sessions – which at the same time is the moment we are interested in the most, we should also investigate other possibilities of gathering more knowledge about a subject's first few minutes with the cross-folding task. These could include the administration of *pretests* which inquire about possible sources of knowledge the subject might employ in their task comprehension (without priming the subject to this knowledge!), and a *prolonged retrospective interview* about these first few minutes (which might however include the danger of retrospective confabulation).

The resolution and detail of our data does not only increase because subjects are gaining practise in reporting. The ease with which one can fashion a report also depends on the task-level knowledge that is (also) being described. In other words: Gathering more knowledge about folding and sketching also makes it easier to talk about it. Since the concepts being described in later sessions tend to be easier to put into words, we could suspect that in the early sessions subjects – still lacking these elaborate concepts – are faced with a *larger amount of verbal overshadowing*. To investigate this possibility, we should look into further possibilities of alternative reporting and avoiding overshadowing (e.g. maybe conduct several short pilots with more elaborate sketching instructions, or working with subjects which are experienced in sketching and drawing).

Another observation in our pilot studies was that given how much possibilities for formalisation and “mathematisation” our task domain allows, subjects could develop far more elaborate ideas about cross-folding if they spend more time with the task. Hence, it would be very interesting to include further sessions at the end of the current ones – increasing the number of folds to be imagined and sketched even further.

A problem that comes along with both, the suggestion of having a second training session and increasing the number of main sessions, is that this would make conducting our study even more demanding (for both, investigators and subjects). Arguably, such an extension would make it even more difficult to find volunteer subjects who would participate in our study even without financial reimbursement (which would be methodologically questionable, cf. Pritchard et al., 1977). Whether it is possible to implement these suggestions thus largely depends on the practical circumstances of our future research.

4.2.2 Further questions

In this last section, we discuss several very broad questions which we think could provide promising guidance for conducting future research based on the work presented.

Informal knowledge and grounding

Our first question is “How could we formally represent knowledge which is not yet formal itself?”. In the preceding chapter we noted that a lot of our subjects’ thinking hinges on metaphors, intuitions and memory episodes and features phenomena such as confusion, mind-wandering and uncertainty. This leads to the question of how we have to represent these kinds of knowledge in our models. How can we formally describe a “hunch” or “vague feeling”? There are philosophical arguments which claim that this endeavour must necessarily fail (cf. e.g. Goel, 1995), but given how relatively little empirical research into thinking about ill-defined problems has been conducted over the last decades, in our opinion we should at least try to gather more data about these kinds of phenomena before giving up on describing them computationally at all.

Another question which follows directly from our first one is “How does this ‘informal’ knowledge develop into formal knowledge?” We observed our subjects starting to think about the cross-folding domain without any domain specific symbolic or syntactic notions, whereas at the end of the sessions all subjects had (to differing extents) developed “language-like” ways of describing folding and sketching. We could also observe that these symbolic notions were grounded in, but not the same as their earlier ways of thinking about the task (e.g. there was a certain amount of transfer of knowledge between the different forms of representation, but it was far from perfect and/or automatic).

While this questions essentially sounds like a concern for developmental psychology (“How does conceptual knowledge develop in principle?”), we want to suggest that it might also be very elucidating to investigate it within adult individuals (“How does person X develop their conceptual knowledge about X from their earlier, non-conceptual understanding?”). In view of our task domain’s propensity to being described mathematically, we could imagine tracing an individual’s path towards a formal apprehension of the task domain, starting from their first intuitive associations when they first read the task instructions⁷.

Introspection and meta-cognition

Even though in this thesis we only touched lightly upon these problems, it was already apparent that there were difficulties in describing the theoretical foundations of our methodology, as well as with the tentative formulation of the meta-cognitive spaces in the preceding chapter. These problems can be

⁷ In some ways, this is what Lakatos described not for individuals, but for the history of mathematics in *Proofs & Refutations* (Lakatos, 1976). In this book, he relates the history of proofs of the Euler characteristic for polyhedra, which started out as informal proofs and thought experiments based on intuition, and were subsequently formalised by an ongoing process of critique and refinement.

seen as symptoms of a general lack of understanding of *meta-cognition*, *introspection* as well as their exact relation. While there is a warranted (if slightly exaggerated) hesitation to employ introspection as an experimental method in cognitive psychology, and while there are many philosophical arguments about introspection as a concept, unfortunately these doubts also lead to hesitation in investigating introspection empirically as a phenomenon of human cognition – sustaining our lack of understanding even further (cf. Jäkel & Schreiber, 2013).

Although there is clear evidence about their positive influences on problem-solving performance (Cox, 2005; Dominowski, 1998; Fox et al., 2011), we do not really know how this influence is being exerted. What does a person actually do when they attend to their own thinking? What are the precise effects of this activity? Can we separate meta-cognition in a broader and introspection in a narrower sense? Can we separate the effects of introspection on its own and having to put introspective observations into words when reporting? The study design presented in this thesis might not necessarily be appropriate to investigate these questions (which might require a more controlled setup), but it is clear that studies like ours would benefit considerably from an increased understanding of these processes.

The role of idiosyncrasies

In presenting our study design, we argued for the importance of conducting single-case studies, because with IDPs we cannot know beforehand how a subject will turn out to represent the task, and which details of their thinking might turn out to be relevant for the subject's individual solution. Yet, although in our task analysis we presented several examples of a subject's *idiosyncrasies*, for the largest part our presentation focussed on abstract commonalities between our pilot subjects, in order to be able to talk about our task in general. It is apparent that had we instead described a single subject's thinking in more detail, we could have provided better explanations of their course of thinking and learning, but fundamentally it is an open question how important such idiosyncrasies are for a theory of problem solving.

To what extent do we need to concern ourselves with them and to what extent can or should we abstract over them (or even ignore them)? Given the theoretical shortcomings of problem space theory we attested in the introduction, we already know that we need to concern ourselves with idiosyncrasies more than it does, but how much detail will turn out to be enough? Could there be a level of detail at which we would merely get lost in irrelevancies which overshadow the possibility of systematisation and generalisation?

Given the (rather low) level of detail we can currently find in our data, for the time being it should be relatively uncontentious to proceed working with the assumption that every detail we can discern might possibly be relevant, but in the future this question will have to be tackled more systematically.

4.3 CONCLUSION

In conclusion, the question of how the cognitive representations a problem solver uses are constructed and change, and whether this activity can be

accurately captured in problem space theory have not yet been answered conclusively. But the thesis at hand contributed to approaching an answer by formulating a set-up and task which allow to investigate representational dynamics in problem solving. Furthermore, our tentative description of a possible multi-problem space architecture to describe subject's thinking when solving our paper-folding task, revealed the difficulties of doing so accurately. These were namely due to the necessity of specifying states and operators to a level of detail beyond what can be found in the empirical data. While on the problem-level (sheets, sketches etc.) there might be enough data to complete these specifications with an acceptable amount of inference and plausible reasoning, there are little to no hints in our data for how we should specify the problem spaces for rule-induction and meta-cognition – and yet these are important constituents of representational dynamics in problem solving.

Having questioned the adequacy of problem spaces for describing these processes, we hence have suggested allow theory formation to consider other kinds of (potentially unconscious) processes, and to focus future research on the questions of how informal knowledge is represented cognitively, how exactly meta-cognitive and introspective processes influence problem solving, and finally to investigate the relevance of the idiosyncrasies of individual thinking with respect to understanding problem solving in general.

BIBLIOGRAPHY

- Agre, P. E. (1988). The Dynamic Structure of Everyday Life. Technical Report 1085, MIT Artificial Intelligence Laboratory. (PhD thesis).
- Agre, P. E. (1994). Interview with Allen Newell. *Artificial Intelligence*, 59:415–449.
- Agre, P. E. & Chapman, D. (1990). What are plans for? *Robotics and Autonomous Systems*, 6(1):17–34.
- Anderson, J. R. & Fincham, J. M. (2014). Extending Problem-solving Procedures through Reflection. *Cognitive Psychology*, 74:1–34.
- Angerer, B. & Schreiber, C. (under review). Investigating Representational Dynamics in Problem Solving. In Olteteanu, A.-M. & Falomir, Z., editors, *Proceedings of the ProSocrates Workshop, Hanse Wissenschaftskolleg, 2017*.
- Anzai, Y. & Simon, H. A. (1979). The Theory of Learning by Doing. *Psychological Review*, 86(2):124.
- Atit, K., Shipley, T. F., & Tikoff, B. (2013). Twisting space: are rigid and non-rigid mental transformations separate spatial skills? *Cognitive Processing*, 14(2):163–173.
- Atwood, M. E., Masson, M. E., & Polson, P. G. (1980). Further explorations with a process model for water jug problems. *Memory & Cognition*, 8(2):182–192.
- Baars, B. J. (1986). *The Cognitive Revolution in Psychology*. New York: The Guilford Press.
- Baars, B. J. & Franklin, S. (2003). How Conscious Experience and Working Memory Interact. *Trends in Cognitive Sciences*, 7(4):166–172.
- Bach, J. (2008). Seven Principles of Synthetic Intelligence. In Wang, P., Goertzel, G., & Franklin, S., editors, *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, pages 63–74. Amsterdam: IOS Press.
- Bach, J. (2012). A framework for emergent emotions, based on motivation and cognitive modulators. *International Journal of Synthetic Emotions (IJSE)*, 3(1):43–63.
- Baddeley, A. (2007). *Working Memory, Thought, and Action*. New York: Oxford University Press.
- Bailes, F., Bishop, L., Stevens, C. J., & Dean, R. T. (2012). Mental Imagery for Musical Changes in Loudness. *Frontiers in Psychology*, 3.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4):723–742.
- Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22:577–660.

- Barth, C. M. & Funke, J. (2010). Negative Affective Environments Improve Complex Solving Performance. *Cognition and Emotion*, 24(7):1259–1268.
- Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and Problem Solving: A Process-Oriented Approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1):205.
- Bickhard, M. H. & Terveen, L. (1996). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. Amsterdam: Elsevier.
- Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence*, 47(1):139–159.
- Bühler, K. (1907). Tatsachen und Probleme zu einer Psychologie der Denkvorgänge I. über Gedanken. *Archiv für die gesamte Psychologie*, 9:297–365.
- Bühler, K. (1908). Tatsachen und Probleme zu einer Psychologie der Denkvorgänge (II. und III.). *Archiv für die gesamte Psychologie*, 12:1–92.
- Burns, B. D. & Vollmeyer, R. (2000). Problem solving: Phenomena in search of a thesis. In Gleitman, L. R. & Joshi, A. K., editors, *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*. Philadelphia, PA: University of Pennsylvania.
- Burns, B. D. & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology: Section A*, 55(1):241–261.
- Chan, J., Paletz, S. B., & Schunn, C. D. (2012). Analogy as a Strategy for Supporting Complex Problem Solving under Uncertainty. *Memory & Cognition*, 40(8):1352–1365.
- Chase, W. G. & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, 4(1):55–81.
- Chi, M. T. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *The Journal of the Learning Sciences*, 6(3):271–315.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5(2):121–152.
- Chi, M. T. & VanLehn, K. A. (2012). Seeing Deep Structure from the Interactions of Surface Features. *Educational Psychologist*, 47(3):177–188.
- Chu, Y. & MacGregor, J. N. (2011). Human Performance on Insight Problem Solving: A Review. *The Journal of Problem Solving*, 3(2):6.
- Cifarelli, V. (1999). Abductive Inference: Connections between Problem Posing and Solving. In Zaslavsky, O., editor, *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 217–224.
- Clancey, W. J. (2006). Observation of Work Practices in Natural Settings. In Ericsson, A., Charness, N., Feltovich, P., & Hoffman, R., editors, *The Cambridge Handbook of Expertise and Expert Performance*, pages 127–145. New York: Cambridge University Press.

- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.
- Clark, A. & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1):7–19.
- Clark, L. V. (1960). Effect of Mental Practice on the Development of a Certain Motor Skill. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 31(4):560–569.
- Clement, J. (2008). *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation*. Springer Science & Business Media.
- Cox, M. T. (2005). Metacognition in computation: A selected research review. *Artificial Intelligence*, 169(2):104–141.
- Croft, D. & Thagard, P. (2002). Dynamic Imagery: A Computational Model of Motion and Visual Analogy. In Magnani, L. & Nersession, N. J., editors, *Model-Based Reasoning*, pages 259–274. Springer.
- Davies, J., Atance, C., & Martin-Ordas, G. (2011). A Framework and Open Questions on Imagination in Adults and Children. *Imagination, Cognition and Personality*, 31(1):143–157.
- de Groot, A. D. (1965). *Thought and Choice in Chess*. Amsterdam University Press.
- Delaney, P. F., Ericsson, K. A., & Knowles, M. E. (2004). Immediate and Sustained Effects of Planning in a Problem-Solving Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6):1219.
- Dietrich, E. & Markman, A. B. (2014). *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*. New York: Psychology Press.
- diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2-3):105–225.
- diSessa, A. A. (2008). How to Observe Knowledge. In *Proceedings of the 8th International Conference on the Learning Sciences*.
- diSessa, A. A. & Cobb, P. (2004). Ontological Innovation and the Role of Theory in Design Experiments. *The Journal of the Learning Sciences*, 13(1):77–103.
- Dominowski, R. L. (1998). Verbalization and Problem Solving. In Hacker, D. L., Dunlosky, J., & Graesser, A., editors, *Metacognition in Educational Theory and Practice*, pages 25–45. Mahwah, NJ: Lawrence Erlbaum.
- Dörner, D. & Kreuzig, H. W. (1983). Problemlösefähigkeit und Intelligenz. *Psychologische Rundschau*, 34:185–192.
- Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Hans Huber.
- Downing, R. E., Moore, J. L., & Brown, S. W. (2005). The Effects and Interaction of Spatial Visualization and Domain Expertise on Information Seeking. *Computers in Human Behavior*, 21(2):195–209.

- Dunbar, K. (1997). How Scientists Think: On-line Creativity and Conceptual Change in Science. In Ward, T., Smith, S., & Vaid, J., editors, *Conceptual Structures and Processes: Emergence, Discovery, and Change*, pages 461–493. Washington, D.C.: APA Press.
- Dunbar, K. & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences*, 5(8):334–339.
- Dunbar, R. I. (2000). Causal Reasoning, Mental Rehearsal, and the Evolution of Primate Cognition. In Heyes, C. & Huber, L., editors, *The Evolution of Cognition*. Cambridge, MA: MIT Press.
- Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Julius Springer.
- Duncker, K. (1945). On Problem-solving. *Psychological Monographs*, 58(5).
- Einstein, A. (1905). On the Electrodynamics of Moving Bodies. *Annal. der Physik*, 17:891–921.
- Ericsson, K. A. (2003). Valid and Non-Reactive Verbalization of Thoughts During Performance of Tasks: Towards a Solution to the Central Problems of Introspection as a Source of Scientific Data. *Journal of Consciousness Studies*, 10(9-10):1–18.
- Ericsson, K. A. & Fox, M. C. (2011). Thinking Aloud Is Not a Form of Introspection but a Qualitatively Different Methodology: Reply to Schooler (2011). *Psychological Bulletin*, 137(2):351–354.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol Analysis – Verbal Reports as Data*. MIT Press, Cambridge, MA. (Revised edition).
- Ericsson, K. A. & Simon, H. A. (1998). How to Study Thinking in Everyday Life: Contrasting Think-aloud Protocols with Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, 5(3):178–186.
- Falomir, Z. (2016). Towards a Qualitative Descriptor for Paper Folding Reasoning. In *Proceedings of the 29th International Workshop on Qualitative Reasoning (QR'16)*, New York City, NY, USA.
- Fazio, L. K. & Siegler, R. S. (2013). Microgenetic learning analysis: A distinction without a difference. *Human Development*, 56(1):52–58.
- Feldon, D. F. (2010). Do psychology researchers tell it like it is? a microgenetic analysis of research strategies and self-report accuracy along a continuum of expertise. *Instructional Science*, 38(4):395–415.
- Finke, R. A. (1993). Chapter 9: Mental Imagery and Creative Discovery. In Roskos-Ewoldsen, B., Intons-Peterson, M. J., & Anderson, R. E., editors, *Imagery, Creativity, and Discovery — A Cognitive Perspective*, volume 98 of *Advances in Psychology*, pages 255–285. Amsterdam: North-Holland.
- Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting Visual Patterns in Mental Imagery. *Cognitive Science*, 13(1):51–78.
- Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. *Journal of Problem Solving*, 4(1):19–42.
- Fischer, A., Holt, D., & Funke, J. (2015). Promoting the Growing Field of Dynamic Decision Making. *Journal of Dynamic Decision Making*, 1(1):1–3.

- Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3-71.
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch Understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*, 3(4):648-666.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin*, 137(2):316-344.
- Frensch, P. A. & Funke, J. (2014). *Complex Problem Solving: The European Perspective*. Psychology Press. (Orig. 1995).
- Friedenberg, J. & Silverman, G. (2011). *Cognitive Science: An Introduction to the Study of Mind*. SAGE Publications. (Second edition).
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer Verlag.
- Funke, J. (2010). Complex Problem Solving: A Case for Complex Cognition? *Cognitive Processing*, 11(2):133-142.
- Funke, J. (2014). Problem solving: What are the important questions? In Bello, P., Guarini, M., McShane, M., & Scassellati, B., editors, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 493-498. Cognitive Science Society.
- Gallivan, B. C. (2002). *How to Fold Paper in Half Twelve Times: An Impossible Challenge Solved and Explained*. Historical Society of Pomona Valley.
- Ginsburg, H. (1981). The Clinical Interview in Psychological Research on Mathematical Thinking: Aims, Rationales, Techniques. *For the Learning of Mathematics*, 1(3):4-11.
- Glaser, B. G. & Strauss, A. L. (2012). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Publishers. (Orig. 1967).
- Glaserfeld, E. v. (1987). Preliminaries to any Theory of Representation. In Janvier, C., editor, *Problems of Representation in the Teaching and Learning of Mathematics*, pages 215-225. Hillsdale, NJ: Erlbaum.
- Goel, V. (1992). "Ill-structured Representations" for Ill-structured Problems. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 130-135. Hillsdale, NJ: Lawrence Erlbaum.
- Goel, V. (1995). *Sketches of Thought*. Cambridge, MA: MIT Press.
- Goel, V. (2010). Neural Basis of Thinking: Laboratory Problems Versus Real-world Problems. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4):613-621.
- Goldstein, E. B. (2011). *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Belmont, CA: Wadsworth. Third edition.
- Gralla, L., Tenbrink, T., Siebers, M., & Schmid, U. (2012). Analogical Problem Solving: Insights from Verbal Reports. In Miyake, N., Peebles, D., & Cooper, R. P., editors, *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pages 396-401. Austin, TX: Cognitive Science Society.

- Granott, N. & Parziale, J. (2002). Microdevelopment: A process-oriented perspective for studying development and learning. In Granott, N. & Parziale, J., editors, *Microdevelopment: Transition processes in development and learning*, pages 1–28. New York: Cambridge University Press.
- Hadamard, J. (1954). *An Essay on the Psychology of Invention in the Mathematical Field*. New York: Dover Publications.
- Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: similarities and differences between mental rotation and mental folding. *Cognitive Processing*, 14(2):105–115.
- Hayes, J. R. & Simon, H. A. (1974). Understanding Written Problem Instructions. In Gregg, L. W., editor, *Knowledge and Cognition*. Hillsdale, NJ: Lawrence Erlbaum.
- Hinton, G. (1979). Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery. *Cognitive Science*, 3(3):231–250.
- Hofstadter, D. R. (2001). Analogy as the Core of Cognition. In Gentner, D., Holyoak, K. J., & Kokinov, B. N., editors, *The Analogical Mind: Perspectives from Cognitive Science*, pages 499–538. Cambridge MA: The MIT Press/Bradford Book.
- Hurley, S. (2001). Perception and Action: Alternative Views. *Synthese*, 129(1):3–40.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Hélie, S. & Sun, R. (2010). Incubation, Insight, and Creative Problem Solving: A Unified Theory and a Connectionist Model. *Psychological Review*, 117(3):994–1024.
- Ida, T., Ghourabi, F., & Takahashi, K. (2015). Formalizing Polygonal Knot Origami. *Journal of Symbolic Computation*, 69:93–108.
- Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology*, 52(6):1122–1131.
- Iyengar, S. R. S., Veni Madhavan, C. E., Zweig, K. A., & Natarajan, A. (2012). Understanding Human Navigation Using Network Analysis. *Topics in Cognitive Science*, 4(1):121–134.
- Jäkel, F. & Schreiber, C. (2013). Introspection in Problem Solving. *The Journal of Problem Solving*, 6(1):4.
- Jeannerod, M. (1995). Mental Imagery in the Motor Context. *Neuropsychologia*, 33(11):1419–1432.
- Karmiloff-Smith, A. (2013). ‘Microgenetics’: No Single Method Can Elucidate Human Learning. *Human Development*, 56(1):47–51.
- Kaufmann, G. (1990). Imagery Effects on Problem Solving. In Hampson, P. J., Marks, D. F., & Richardson, J. T. E., editors, *Imagery: Current Developments*, pages 169–196. Florence, KY, US: Taylor & Frances/Routledge.
- Kirsh, D. (2009a). Problem solving and situated cognition. In Robbins, P. & Aydede, M., editors, *The Cambridge Handbook of Situated Cognition*, pages 264–306. Cambridge University Press.

- Kirsh, D. (2009b). Projection, Problem Space and Anchoring. In Taatgen, N. & van Rijn, H., editors, *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Kistner, S., Burns, B. D., Vollmeyer, R., & Kortenkamp, U. (2014). An Explorative Study of Search of Model Space in Problem-solving. *Journal of Cognitive Psychology*, 26(7):818–829.
- Klahr, D. & Dunbar, K. (1988). Dual Space Search during Scientific Reasoning. *Cognitive Science*, 12(1):1–48.
- Kluwe, R. H. (1995). Single Case Studies and Models of Complex Problem Solving. In Frensch, P. A. & Funke, J., editors, *Complex Problem Solving: The European Perspective*, pages 269–291. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Knoblich, G., Ohlsson, S., Haider, S., & Rhenius, D. (1999). Constraint Relaxation and Chunk Decomposition in Insight Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25:1534–1555.
- Knuth, D. E. (1978). Lewis Carroll's Word, Ward, Ware, Dare, Dame, Game. *GAMES*, 2(4):22–23.
- Koedinger, K. R. & Anderson, J. R. (1990). Abstract Planning and Perceptual Chunks: Elements of Expertise in Geometry. *Cognitive Science*, 14(4):511–550.
- Kosslyn, S. M. (1996). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the Demystification of Mental Imagery. *Behavioral and Brain Sciences*, 2(4):535–548.
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*, volume vi. New York: Oxford University Press.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why Are Some Problems Hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2):248–294.
- Krumnack, U., Schwering, A., Kühnberger, K.-U., Gust, H., Abdel-Fattah, A., Besold, T., Schmidt, M., & Schneider, S. (2013). Sketch Learning by Analogy. In Kutz, O., Bhatt, M., Borgo, S., & Santos, P., editors, *Proceedings of the Workshop SHAPES 2.0, Co-located with UNI-LOG 2013, Rio de Janeiro, Brazil*.
- Köhler, W. (1947). *Gestalt Psychology*. New York: Liveright Publishing Corporation.
- Külpe, O. (1912). Über die moderne Psychologie des Denkens. *Internationale Monatsschrift für Wissenschaft, Kunst und Technik*, 6:1069–1110.
- Külpe, O. (1920). *Vorlesungen über Psychologie*. Leipzig: G. Hirzel. (published posthumously by K. Bühler).
- Laird, J. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.
- Laird, J., Rosenbloom, P., & Newell, A. (1986). *Universal Subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies*. Kluwer Academic Publishers.

- Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press.
- Larkin, J. H. & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, 11(1):65–100.
- Leahy, W. & Sweller, J. (2008). The Imagination Effect Increases with an Increased Intrinsic Cognitive Load. *Applied cognitive psychology*, 22(2):273–283.
- León-Villagrà, P. & Jäkel, F. (2013). Categorization and Abstract Similarity in Chess. In Knauff, M., Pauen, M., Sebanz, N., & Wachsmuth, I., editors, *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pages 2860–2865. Austin, TX: Cognitive Science Society.
- Liepmann, H. (1904). *Über Ideenflucht: Begriffsbestimmung und psychologische Analyse*. C. Marhold.
- Lucas, E. (1979). *Récréations mathématiques*, Vol. 3 (1891–1894). Paris: Gauthier-Villars.
- Machón, A. (2013). *Children's Drawings: The Genesis and Nature of Graphic Representation: A Developmental Study*. Madrid, Spain: Fíbulas.
- Maier, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, 12:181–194.
- Mandler, G. (2011). *A History of Modern Experimental Psychology: From James and Wundt to Cognitive Science*. Cambridge, MA: MIT Press.
- Mandler, J. M. & Mandler, G. (1964). *Thinking: From Association to Gestalt*. New York: Wiley.
- Markman, A. B. & Dietrich, E. (2000). Extending the Classical View of Representation. *Trends in Cognitive Sciences*, 4(12):470–475.
- Marr, D. & Poggio, T. (1976). From Understanding Computation to Understanding Neural Circuitry. AI Memo 357, MIT.
- Michie, D. (1968). Memo Functions and Machine Learning. *Nature*, 218:19–22.
- Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE*, 49(1):8–30.
- Minsky, M. (1991). Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy. *AI Magazine*, 12(2):34–51.
- Minsky, M. (1992). Future of AI Technology. *Toshiba Review*, 47(7). Retrieved from <http://web.media.mit.edu/~minsky/papers/CausalDiversity.html>.
- Minsky, M. (2006). *The Emotion Machine*. New York: Simon and Schuster.
- Newell, A. (1966). On the Analysis of Human Problem Solving Protocols. Technical Report 1942, Computer Science Department, Carnegie Mellon University.
- Newell, A. (1967). Studies in Problem Solving; Subject 3 on the Cryptarithmic Task Donald+Gerald=Robert. Technical Report 1731, Computer Science Department, Carnegie Mellon University.

- Newell, A. (1979). Reasoning, Problem Solving and Decision Processes: The Problem Space as a Fundamental Category. Technical Report 2482, Computer Science Department, Carnegie Mellon University.
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a Theory of Human Problem Solving. *Psychological Review*, 65(3):151.
- Newell, A., Shaw, J. C., & Simon, H. A. (1959). A Variety of Intelligent Learning in a General Problem Solver. Technical Report P-1742, Rand Corporation.
- Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Prentice-Hall Englewood Cliffs, NJ.
- Newell, A. & Simon, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 19(3):113–126.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3):231.
- Nokes, T., Schunn, C., & Chi, M. (2010). Problem Solving and Human Expertise. In Peterson, P., Baker, E., & McGaw, B., editors, *International Encyclopedia of Education*, volume 5, pages 265–272. Elsevier.
- Noyes, J. M. & Garland, K. J. (2003). Solving the Tower of Hanoi: does mode of presentation matter? *Computers in Human Behavior*, 19(5):579–592.
- Noë, A. & O'Regan, K. J. (2000). Perception, Attention, and the Grand Illusion. *Psyche*, 6(15).
- OEIS Foundation Inc. (2016). The On-Line Encyclopedia of Integer Sequences. published electronically at <https://oeis.org>. Sequence A027383.
- Ohlsson, S. (1990). Trace Analysis and Spatial Reasoning: An Example of Intensive Cognitive Diagnosis and Its Implications for Testing. *Diagnostic Monitoring of Skill and Knowledge Acquisition*, pages 251–296.
- Ohlsson, S. (2008). Computational Models of Skill Acquisition. In Sun, R., editor, *The Cambridge Handbook of Computational Psychology*, chapter 13, pages 359–395. Cambridge University Press.
- Ohlsson, S. (2010). Questions, Patterns and Explanations, Not Hypothesis Testing, is the Core of Psychology as of Any Science. In Toomela, A. & Valsiner, J., editors, *Methodological Thinking in Psychology: 60 Years Gone Astray?* Charlotte, NC: Information Age Publishing.
- Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. New York: Cambridge University Press.
- Ohlsson, S. (2012a). Cognitive Skill Acquisition. In Seel, N. M., editor, *Encyclopedia of the Sciences of Learning*, pages 616–619. New York: Springer. DOI: 10.1007/978-1-4419-1428-6_297.
- Ohlsson, S. (2012b). The Problems with Problem Solving: Reflections on the Rise, Current Status, and Possible Future of a Cognitive Research Paradigm. *The Journal of Problem Solving*, 5(1):7.

- Overgaard, M., Gallagher, S., & Ramsøy, T. Z. (2008). An integration of first-person methodologies in cognitive science. *Journal of Consciousness Studies*, 15(5):100–120.
- Pang, W., Wang, K.-P., Zhou, C.-G., Dong, L.-J., Liu, M., Zhang, H.-Y., & Wang, J.-Y. (2004). Modified Particle Swarm Optimization Based on Space Transformation for Solving Traveling Salesman Problem. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*. Institute of Electrical and Electronics Engineers (IEEE).
- Parnafes, O. & diSessa, A. A. (2013). Microgenetic Learning Analysis: A Methodology for Studying Knowledge in Transition. *Human Development*, 56(1):5–37.
- Pearson, J. & Kosslyn, S. M. (2013). Mental Imagery. *Frontiers in Psychology*, 4.
- Petitmengin, C. (2006). Describing One's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness. *Phenomenology and the Cognitive Sciences*, 5(3-4):229–269.
- Poincaré, H. (1913). *The Foundations of Science*. New York: The Science Press. Authorised translation by G. B. Halsted.
- Pritchard, R. D., Campbell, K. M., & Campbell, D. J. (1977). Effects of Extrinsic Financial Rewards on Intrinsic Motivation. *Journal of Applied Psychology*, 62(1):9–15.
- Pylyshyn, Z. W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, 80(1):1–24.
- Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. A Bradford Book.
- Pólya, G. (1945). *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press.
- Qin, Y. & Simon, H. A. (1990). Laboratory replication of scientific discovery processes. *Cognitive Science*, 14(2):281–312.
- Qin, Y. & Simon, H. A. (1992a). Imagery and Mental Models in Problem Solving. In *Proc. AAAI Symposium on Reasoning with Diagrammatic Representations*, Stanford, CA.
- Qin, Y. & Simon, H. A. (1992b). Imagery as Process Representation in Problem Solving. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 1050–1055.
- Quesada, J., Kintsch, W., & Gomez, E. (2005). Complex Problem-solving: A Field in Search of a Definition? *Theoretical Issues in Ergonomics Science*, 6(1):5–33.
- Raab, T. (2015a). Selbstbeobachtung. Wozu und, wenn ja, welche? Eine Einleitung zur Denkpsychologie Oswald Wieners. In Raab, T. & Eder, T., editors, *Selbstbeobachtung: Oswald Wieners Denkpsychologie*, pages 9–56. Berlin: Suhrkamp.
- Raab, T. (2015b). Vermessung | Selbstbeobachtung zu einem Diktum von Helmholtz. In Raab, T. & Eder, T., editors, *Selbstbeobachtung: Oswald Wieners Denkpsychologie*, pages 217–236. Berlin: Suhrkamp.

- Raab, T. & Eder, T., editors (2015). *Selbstbeobachtung: Oswald Wieners Denkpsychologie*. Berlin: Suhrkamp.
- Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. New York: Oxford University Press.
- Reed, S. K. (1993). Chapter 10: Imagery and Discovery. In Roskos-Ewoldsen, B., Intons-Peterson, M. J., & Anderson, R. E., editors, *Imagery, Creativity, and Discovery — A Cognitive Perspective*, volume 98 of *Advances in Psychology*, pages 287–312. Amsterdam: North-Holland.
- Reed, S. K. (2015). The Structure of Ill-Structured (and Well-Structured) Problems Revisited. *Educational Psychology Review*, pages 1–26.
- Reitman, W. R. (1965). *Cognition and Thought: An Information Processing Approach*. New York: Wiley.
- Ruiz, D. & Newell, A. (1989). Tower-noticing Triggers Strategy-change in the Tower of Hanoi: A Soar Model. Technical Report – AIP 66, Carnegie Mellon University.
- Scherbel, M. (1999). *Phänomenologie als absolute Wissenschaft: die systembildende Funktion des Zuschauers in Eugen Finks VI. Cartesianischer Meditation*, volume 75 of *Elementa*. Amsterdam: Rodopi.
- Schmid, U. (2005). A Cognitive Model of Learning by Doing. In Bab, S., Gulden, J., Noll, T., & Wiczorek, T., editors, *Models and Human Reasoning – Bernd Mahr zum 60. Geburtstag*, pages 235–252. Wissenschaft und Technik Verlag, Berlin.
- Schneider, S., Krumnack, U., & Kühnberger, K.-U. (2013). Dynamic Assembly of Figures in Visuospatial Reasoning. In Kutz, O., Bhatt, M., Borgo, S., & Santos, P., editors, *Proceedings of the Workshop SHAPES 2.0, Co-located with UNI-LOG 2013, Rio de Janeiro, Brazil*.
- Schoenfeld, A. H., Smith, J. P., & Arcavi, A. (1993). Learning: The Microgenetic Analysis of One Student's Evolving Understanding of a Subject Matter Domain. In Glaser, R., editor, *Advances in Instructional Psychology*, volume 4, pages 55–175. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schooler, J. W. & Engstler-Schooler, T. Y. (1990). Verbal Overshadowing of Visual Memories: Some Things Are Better Left Unsaid. *Cognitive Psychology*, 22(1):36–71.
- Schreiber, C. (2015). Topologie | Konstruktion einer Kreuzfaltung: Selbstbeobachtungen zum lebendigen Denken. In Raab, T. & Eder, T., editors, *Selbstbeobachtung: Oswald Wieners Denkpsychologie*, pages 191–216. Berlin: Suhrkamp.
- Schunn, C. D. & Dunbar, K. (1996). Priming, Analogy, and Awareness in Complex Reasoning. *Memory & Cognition*, 24(3):271–284.
- Schunn, C. D. & Klahr, D. (1995). A 4-space Model of Scientific Discovery. In Moore, J. D. & Lehman, J. F., editors, *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 106–111.
- Schunn, C. D. & Klahr, D. (1996). The problem of problem spaces: When and how to go beyond a 2-space model of scientific discovery. In Cottrell, G. W., editor, *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pages 25–26.

- Schwank, I. (1999). On Predicative versus Functional Cognitive Structures. In Schwank, I., editor, *Proceedings of the First Conference of the European Society in Mathematics Education Vol. II*, pages 84–96. Forschungsinstitut fuer Mathematikdidaktik, Osnabrück.
- Schwartz, S. M. & Fattaleh, D. L. (1972). Representation in Deductive Problem-solving: The Matrix. *Journal of Experimental Psychology*, 95(2):343–348.
- Selz, O. (1913). Über die Gesetze des geordneten Denkverlaufs. *Stuttgart: Spemann*.
- Selz, O. (1927). Die Umgestaltung der Grundanschauungen vom intellektuellen Geschehen. *Kant-Studien*, 32:273–280.
- Shepard, R. N. & Feng, C. (1972). A Chronometric Study of Mental Paper Folding. *Cognitive Psychology*, 3(2):228–243.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703.
- Siegler, R. S. (2002). Microgenetic Studies of Self-explanation. In Granott, N. & Parziale, J., editors, *Microdevelopment: Transition processes in development and learning*, pages 31–58. New York: Cambridge University Press.
- Siegler, R. S. (2006). Microgenetic Analyses of Learning. In Kuhn, D. & Siegler, R. S., editors, *Handbook of Child Psychology (6th edition)*, volume 2: Cognition, Perception, and Language, pages 464–510. New York: Wiley.
- Siegler, R. S. & Chen, Z. (1998). Developmental Differences in Rule Learning: A Microgenetic Analysis. *Cognitive Psychology*, 36(3):273–310.
- Siegler, R. S. & Crowley, K. (1991). The Microgenetic Method: A Direct Means for Studying Cognitive Development. *American Psychologist*, 46(6):606–620.
- Siegler, R. S. & Stern, E. (1998). Conscious and Unconscious Strategy Discoveries: A Microgenetic Analysis. *Journal of Experimental Psychology: General*, 127(4):377–397.
- Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, 74(1):29.
- Simon, H. A. (1973). The Structure of Ill-Structured Problems. *Artificial Intelligence*, 4(3-4):181–201.
- Simon, H. A. (1991). Cognitive Architectures and Rational Analysis: Comment. In VanLehn, K., editor, *Architectures for Intelligence: The 22nd Carnegie Mellon Symposium on Cognition*, pages 25–39. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Simon, H. A. & Hayes, J. R. (1976). The Understanding Process: Problem Isomorphs. *Cognitive Psychology*, 8(2):165–190.
- Simon, H. A. & Lea, G. (1974). Problem Solving and Rule Induction: A Unified View. In Gregg, L. W., editor, *Knowledge and Cognition*, pages 105–127. Hillsdale, NJ: Erlbaum.
- Simon, H. A. & Newell, A. (1958). Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research*, 6(1):1–10.

- Simon, H. A. & Newell, A. (1971). Human Problem Solving: The State of the Theory in 1970. *American Psychologist*, 26(2):145–159.
- Simon, H. A. (with Wiener, O.) (1994). Kambrium der Künstlichen Intelligenz. In *Die Wissenschaft vom Künstlichen*, pages 175–228. Wien: Springer. (Afterword by Wiener).
- Someren, M. W. v., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Approach to Modelling Cognitive Processes*. London, UK: Academic Press.
- Stenning, K. & van Lambalgen, M. (2004). A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, 28(4):481–529.
- Sun, R., Slusarz, P., & Terry, C. (2005). The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach. *Psychological Review*, 112(1):159–192.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York: Springer.
- Taylor, H. A. & Tenbrink, T. (2013). The Spatial Thinking of Origami: Evidence from Think-aloud Protocols. *Cognitive Processing*, 14(2):189–191.
- Thagard, P. (2005). *Mind: Introduction to Cognitive Science*. Cambridge, MA: MIT Press. Second edition.
- Tschirgi, J. E. (1980). Sensible Reasoning: A Hypothesis about Hypotheses. *Child Development*, 51(1):1–10.
- Tversky, B. (2005). Visuospatial Reasoning. In Holyoak, K. J. & Morrison, R. G., editors, *The Cambridge Handbook of Thinking and Reasoning*, pages 209–240. Cambridge, MA: Cambridge University Press.
- Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press.
- Van der Waerden, B. (1953). Einfall und Überlegung in der Mathematik. *Elemente der Mathematik*, 8:121–129.
- VanLehn, K. (1988). Problem Solving and Cognitive Skill Acquisition. Tech report – AIP 32, Carnegie Mellon University.
- VanLehn, K. (1991). Rule Acquisition Events in the Discovery of Problem-Solving Strategies. *Cognitive Science*, 15(1):1–47.
- VanLehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology*, 47(1):513–539.
- VanLehn, K. & Ball, W. (1991). Goal Reconstruction: How Teton Blends Situated Action and Planned Action. In VanLehn, K., editor, *Architectures for Intelligence: The 22nd Carnegie Mellon Symposium on Cognition*, pages 147–188. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- VanLehn, K., Brown, J. S., & Greeno, J. (1982). Competitive Argumentation in Computational Theories of Cognition. CIS 14, Cognitive and Instructional Sciences Group, Xerox Corporation, Palo Alto Research Center.

- Vendetti, M. S., Wu, A., Rowshanshad, E., Knowlton, B. J., & Holyoak, K. J. (2014). When Reasoning Modifies Memory: Schematic Assimilation Triggered by Analogical Mapping. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(4).
- Voneche, J. (1999). The Origin of Piaget's Ideas about Genesis and Development. In Scholnick, E. K., Nelson, K. K., Gelman, S. A., & Miller, P. H., editors, *Conceptual Development. Piaget's Legacy*, pages 243–252. Mahwah, NJ: Lawrence Erlbaum.
- Weisberg, R. & Suls, J. M. (1973). An Information-Processing Model of Duncker's Candle Problem. *Cognitive Psychology*, 4(2):255–276.
- Wenke, D. & Frensch, P. A. (2003). Is Success or Failure at Solving Complex Problems Related to Intellectual Ability? In Davidson, J. E. & Sternberg, R. J., editors, *The Psychology of Problem Solving*, pages 87–126. New York: Cambridge University Press.
- Wertheimer, M. (1959). *Productive Thinking*. New York: Harper & Row.
- Wiener, O. (2000). *Materialien zu meinem Buch VORSTELLUNGEN*, volume 05 of *AUSSCHNITT*. Wien: Technische Universität.
- Wiener, O. (2006). Humbug. *Der Ficker*, 2:96–116.
- Wiener, O. (2008). Über das "Sehen" im Traum (Dritter Teil). *manuskripte*, 181:132–141.
- Wiener, O. (2015). Kybernetik und Gespenster. *manuskripte*, pages 143–163.
- Winograd, T. (1972). Understanding Natural Language. *Cognitive Psychology*, 3(1):1–191.
- Wintermute, S. (2009). An Overview of Spatial Processing in Soar/SVS. Technical Report CCA-TR-2009-01, Center for Cognitive Architecture, University of Michigan.
- Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training Generalized Spatial Skills. *Psychonomic Bulletin & Review*, 15(4):763–771.

APPENDIX A: STUDY MATERIAL

A.1 CONSENT FORM

Der/die Versuchsleiter/-in hat mich hinreichend über Ablauf und Form des betreffenden Experiments, insbesondere die Freiwilligkeit meiner Teilnahme, informiert und ich erkläre mich hiermit einverstanden, dass die im Rahmen des Experiments erhobenen Daten und daraus hervorgehende Untersuchungsergebnisse in anonymisierter Form zu Forschungszwecken verwendet werden.

Datum,

(Unterschrift der/des Versuchsteilnehmer/-in)
(Unterschrift der/des Versuchsleiter/-in)

A.2 GENERAL INSTRUCTIONS

Allgemeine Instruktionen:

Die Aufgaben, die wir Dir stellen werden, sind keine Leistungstests, es werden keinerlei „Performance“-Messungen von uns vorgenommen, d. h. es geht nicht darum, möglichst schnell zu antworten, sondern es kommt uns darauf an, dass Du uns berichtest, *wie* sich Dein Bearbeiten der Aufgabe gestaltet.

Wir möchten Dich bitten, alle Aufgaben, die wir Dir stellen, allein in Deiner *Vorstellung*, also ohne externe Hilfsmittel zu bearbeiten. Ob Du dazu die Augen schließt, kannst Du selbst entscheiden.

Des Weiteren möchten wir Dich bitten, beim Bearbeiten der Aufgaben eine „*selbstbeobachtende Haltung*“ einzunehmen, d. h. Dein eigenes Denken aufmerksam mitzuverfolgen. Dabei kann es auch geschehen, dass Dir sinnesartige, gefühlsmäßige oder körperliche Aspekte deines Denkens auffallen.

Beim *Berichten* deiner Beobachtungen ist uns wichtiger, dass Du eher den Verlauf Deiner Gedanken schilderst, als möglichst genau und eloquent deren Inhalte zu schildern. Für den Fall, dass Du Teilverläufe nicht berichten möchtest, möchten wir Dich bitten, solche Auslassungen an entsprechender Stelle kurz zu vermerken.

Im Folgenden schildern wir Dir einige bekannte „*Stolperfallen*“, die beim Berichten von Selbstbeobachtungen passieren können, um zu vermeiden, dass du in sie „tappst“:

Auslassungen: Wie quasi jede Methode zur Datenerfassung kann auch die Selbstbeobachtung nicht alles erfassen, was im Nachhinein vielleicht relevant sein könnte. Lass Dich also nicht frustrieren, wenn Dir eine Beobachtung entflieht – solange Du allerdings noch sagen kannst, dass „etwas“ passiert ist, solltest Du es aber auch berichten.

Inädaquate Bezeichner: Die andere Seite desselben Problems ist, dass man manchmal geneigt ist, einer Beobachtung zu selbstsicher einen Namen/ein Konzept/eine Erscheinung/... zuzuordnen, auch wenn man schon merkt, dass es eher „nur so ungefähr“ stimmt. Hier gilt genauso: Lieber die eigene Unklarheit mit einem „irgendwie“ o. ä. markieren, als sich zu schnell zu weit aus dem Fenster zu lehnen.

Sprachliche Übershattungen: Manchmal mag es passieren, dass in der Vorstellung Dinge geschehen, die sich nicht gut in Worte fassen lassen. Für diesen Fall stellen wir Dir Stift und Papier zur Verfügung, so dass Du ggf. kleine Skizzen anfertigen kannst.

Erklären/Räsonieren: Gerade Personen mit einem eigenen, starken Interesse an Denkvorgängen geraten beim Selbstbeobachten manchmal schnell in eine spekulative Haltung, in der sie mit dem Bericht darüber, was geschehen ist auch gleich hypothetisieren, *warum* das geschehen sein könnte. Solltest Du bemerken, dass Du ins Spekulieren gerätst, bitten wir Dich, Dich wieder auf die Aufgabe zu besinnen und ggf. Stellen im Bericht zu markieren, von denen Du meinst, dass sie spekulativer Natur sind.

A.3 PROTOCOL INSTRUCTIONS

Instruktionen zur selbstständigen Protokollerstellung:

Für die folgende Aufgabe bitten wir Dich, *selbstständig* ein schriftliches Protokoll Deiner Selbstbeobachtungen zu erstellen. Was das genaue Vorgehen angeht, muss jeder seinen eigenen Stil finden. Nichtsdestotrotz sollen Dir die folgenden Hinweise helfen zu vermeiden, dass sich Bearbeiten der Aufgabe und Verfassen von Notizen zu sehr in die Quere kommen.

Versuche es zu vermeiden, gleichzeitig die Aufgabe zu bearbeiten und Notizen anzufertigen. Anstatt dessen unterbreche Dein Bearbeiten der Aufgabe hin und wieder und halte rückblickend Deine Selbstbeobachtungen fest. Versuche Dich dabei so kurz wie möglich und so genau wie nötig auszudrücken.

Zur Unterbrechung ereignen sich meist von selbst gewisse „Sollbruchstellen“ im Denken, z.B. zu Ende geführte Teilaufgaben oder Sackgassen, in denen man Notizen anfertigen kann.

Des Weiteren kann es passieren, dass besonders eindruckliche oder auch als flüchtig empfundene Beobachtungen gemacht werden. Auch in diesen Fällen solltest Du Dein Bearbeiten der Aufgabe kurz pausieren und eine Notiz machen.

Im Allgemeinen solltest Du Deine Beschäftigung mit der Aufgabe also nur dann zum Zweck des Dokumentierens unterbrechen, wenn Du selbst das Gefühl hast, dass die Unterbrechung nicht weiter störend ist oder wenn Du etwas Wichtiges zu notieren meinst, das Dir sonst entkommen würde.

Falls Dir die sprachliche Beschreibung Schwierigkeiten bereiten sollte, hast Du zudem die Möglichkeit auf Papier kleine Skizzen anzufertigen, die Deine Selbstbeobachtungen dokumentieren (Bitte nummeriere die Skizzen und verwende entsprechende Bezeichner in deinem schriftlichen Protokoll, so dass im Nachhinein zuortbar ist, welche Skizze wann entstanden ist).

Bedenke jedoch, dass du es beim Skizzieren nach Möglichkeit vermeiden solltest, die Aufgabe weiterzubearbeiten oder die Skizze als „Denkwerkzeug“ zu benutzen. Um das zu vermeiden, drehe das Blatt mit der Skizze bitte gleich nach dem Zeichnen um. Sollte es Dir dennoch einmal passieren, dass eine Skizze, die Du angefertigt hast, Dich auf eine neue Idee bringt, bitten wir Dich das im Bericht festzuhalten.

A.4 SHORTENED INSTRUCTIONS (REMINDER)

Zusammenfassung der Instruktionen:

Die vorliegenden Aufgaben sind kein Leistungstest. Manche von ihnen sind vergleichsweise schwierig. Lass Dir also Zeit beim Denken und lass Dich nicht von anfänglicher Verwirrung beunruhigen.

Bearbeite die Aufgaben *allein in der Vorstellung* und mit *selbstbeobachtender Haltung* (achte ggf. auf sinnesartige, gefühlsbezogene oder körperliche Aspekte).

Trenne das Bearbeiten der Aufgabe vom Verfassen von Notizen. Um das zu vereinfachen, lass Deine Hände bitte nicht auf der Tastatur ruhen. Vergiss allerdings nicht, zu protokollieren.

Unterbreche das Bearbeiten der Aufgabe ereignisbedingt, z.B. nach Beendigung von Teilaufgaben, in Sackgassen sowie bei bemerkenswerten oder flüchtigen Ereignissen.

Verfasse während der Unterbrechungen retrospektiv schriftliche Notizen („so kurz wie möglich und so genau wie nötig“) oder Skizzen über den Verlauf Deiner Gedanken.

Verwende bitte weder Deine bisherigen Verschriftlichungen noch Deine eventuellen Skizzen als Denkwerkzeug. Markiere im Protokoll, falls es dennoch passiert.

Beachte beim Verfassen von Notizen den Umgang mit den typischen Fehlerquellen der Selbstbeobachtung, z.B. Auslassungen, inadäquate Bezeichner, sprachliche Überschattungen und spekulatives Rasonieren.

A.5 TASKS

A.5.1 *Training session*

First block:

1. Stelle Dir bitte ein großes H vor.
2. Jetzt, rotiere es bitte um 90° nach rechts.
3. Jetzt, füge oben ein Dreieck an, dessen Basis gleich breit ist.
4. Kannst Du die entstandene Figur erkennen?

(Baum)

1. Stelle Dir bitte ein großes D vor.
 2. Jetzt, rotiere es bitte um 90° nach links.
 3. Jetzt, füge bitte ein großes J an die Unterseite der Figur.
 4. Kannst Du die entstandene Figur erkennen?
- (Regenschirm)

1. Stelle Dir bitte ein großes T vor.
 2. Jetzt, rotiere es bitte um 180° .
 3. Jetzt, füge oben bitte ein Dreieck an, so dass seine Basis oben ist und es nach unten zeigt.
 4. Kannst Du die entstandene Figur erkennen?
- (Martiniglas bzw. Straßenschild)

Second block:

1. Stelle Dir ein Quadrat vor.
 2. Teile die beiden Senkrechten Seiten in je drei gleiche Teile.
 3. Verbinde die so entstandenen Punkte durch waagrechte Geraden, so, dass drei gleich große Rechtecke entstehen.
 4. Nun zeichne die beiden Diagonalen des Quadrats ein.
 5. Wieviele Dreiecke enthält diese Konstruktion?
- (18)

A.5.2 Session 1

First block

The first task block of session 1 consists of “warm-up tasks”, which were selected such that they prepare subjects for (a) solving transformational problems, and (b) manipulate complex mental imagery in a familiar situative context.

The task block includes four tasks. The first three are *word-morph* problems. Word-morph is a simple game, where the goal is to link two words with a chain of words where only one letter at a time is allowed to change, e.g. WEB $\rightarrow$ BIN, could be connected with the sequence WED, BED, BID, BIN (cf. Knuth, 1978; Iyengar et al., 2012).

The selected problems were:

- WEB $\rightarrow$ BIN,
- ARE $\rightarrow$ WIG,
- ODD $\rightarrow$ GYM

If the investigator should notice that the subject has difficulties with the task (e.g. because they know too few English three-letter words), then one or even two of these tasks may be skipped. The upper time limit for the first

task should be around 10 minutes, the two subsequent ones should not last longer than 5 minutes.

The original task instruction is:

Gegeben sind zwei englische Wörter. Ziel der Aufgabe ist es, eine Kette englischer Wörter gleicher Buchstabenanzahl zu finden, ausgehend vom einen, hin zum anderen gegebenen Wort, so dass in der Kette aufeinander folgende Wörter sich jeweils in nur einem Buchstaben unterscheiden.

The final warm-up tasks consists solely of the instruction:

Versuche Dir bitte vorzustellen, wie du deine Schuhe bindest und verfasse eine kurze, schrittweise Anweisung für das Knüpfen der Schuhmasche, wie du sie üblicherweise machst.

(Please imagine how you tie your shoe laces and give a short, step-wise explanation of how you make the knot you are usually using.)

This task should not last longer than 20 minutes. After quarter of an hour, the subject should be asked to slowly finish the task, or if they feel they can not, to try reporting about their difficulty of imagining the knot.

Second block

T1.2 is displayed over two slides (as are T2.1, T2.2, T3.1, T3.2, T4.1, T4.2, which all display the following introductory slide before their actual task instruction):

Slide 1:

Bitte lies Dir die folgende Aufgabe genau durch und präge Dir die Aufgabenstellung ein. Wenn Du weißt, was zu tun ist, klicke bitte auf den Button „Zur Aufgabe“. Die auf den Aufgabentext folgende Seite ist absichtlich leer, um Dir Raum zu geben Dich auf das Bearbeiten der Aufgabe und das Verfassen Deiner Notizen zu konzentrieren.

Wenn etwas unklar sein sollte, steht Dir der Versuchsleiter gerne zur Verfügung.

Slide 2:

Bitte bearbeite die Aufgabe solange, bis du selbst meinst, sie gut genug zu beherrschen, um einen dazugehörigen Test bestehen zu können. Wenn es so weit ist, kannst du selbstständig zum Test fortfahren. Solltest du allerdings feststellen, dass du den Test nicht ohne weiteres durchführen kannst, sondern noch weiter vorstellen musst, bitten wir Dich dies zu tun und deine Weiterarbeit daran wie zuvor zu protokollieren. Den Test solltest du also erst bearbeiten, wenn du dafür keine nennenswerte Denkarbeit mehr leisten musst.

Die Aufgabe:

Falte ein Blatt in deiner Vorstellung zweimal nacheinander, sodass es durch jede Faltung mittig halbiert wird, und die entstehenden Faltkanten senkrecht zueinander stehen.

Mach dich mit dem Gegenstand nach jeder Faltung vertraut.

T1.3:

Bitte stelle in je einer schematischen Skizze das zweifach gefaltete Blatt von links, rechts, oben, unten, hinten und vorne dar (insgesamt also 6 2D-Skizzen). Markiere unter der jeweiligen Skizze bitte auch, welche Ansicht sie darstellt.

Zeichne bitte erst dann, wenn Du Dir über jeden Strich, den Du machen willst, in der Vorstellung bereits sicher bist.

A.5.3 Session 2

First block

T2.1 is similar to T1.2 except that it does not include the announcement of the follow-up test (T1.3)

Second block

T2.2:

Bitte bearbeite die Aufgabe solange, bis du selbst meinst, sie gut genug zu beherrschen, um einen dazugehörigen Test bestehen zu können. Wenn es so weit ist, kannst du selbstständig zum Test fortfahren. Solltest du allerdings feststellen, dass du den Test nicht ohne weiteres durchführen kannst, sondern noch weiter vorstellen musst, bitten wir Dich dies zu tun und deine Weiterarbeit daran wie zuvor zu protokollieren. Den Test solltest du also erst bearbeiten, wenn du dafür keine nennenswerte Denkarbeit mehr leisten musst.

Die Aufgabe:

*Falte ein Blatt in deiner Vorstellung **dreimal** nacheinander, sodass es durch jede Faltung mittig halbiert wird, und die entstehenden Faltkanten senkrecht zueinander stehen.*

Mach dich mit dem Gegenstand nach jeder Faltung vertraut.

T2.3:

Bitte stelle in je einer schematischen Skizze das **dreifach** gefaltete Blatt von links, rechts, oben, unten, hinten und vorne dar (insgesamt also 6 2D-Skizzen). Markiere unter der jeweiligen Skizze bitte auch, welche Ansicht sie darstellt.

Zeichne bitte erst dann, wenn Du Dir über jeden Strich, den Du machen willst, in der Vorstellung bereits sicher bist.

A.5.4 Session 3

First block

T3.1 is similar to T2.2 except that it does not include the announcement of the follow-up test (T2.3)

Second block

T3.2:

Bitte bearbeite die Aufgabe solange, bis du selbst meinst, sie gut genug zu beherrschen, um einen dazugehörigen Test bestehen zu können. Wenn es so weit ist, kannst du selbstständig zum Test fortfahren. Solltest du allerdings feststellen, dass du den Test nicht ohne weiteres durchführen kannst, sondern noch weiter vorstellen musst, bitten wir Dich dies zu tun und deine Weiterarbeit daran wie zuvor zu protokollieren. Den Test solltest du also erst bearbeiten, wenn du dafür keine nennenswerte Denkarbeit mehr leisten musst.

Die Aufgabe:

Falte ein Blatt in deiner Vorstellung **viermal** nacheinander, sodass es durch jede Faltung mittig halbiert wird, und die entstehenden Faltkanten senkrecht zueinander stehen.

Mach dich mit dem Gegenstand nach jeder Faltung vertraut.

T3.3:

Bitte stelle in je einer schematischen Skizze das **vierfach** gefaltete Blatt von links, rechts, oben, unten, hinten und vorne dar (insgesamt also 6 2D-Skizzen). Markiere unter der jeweiligen Skizze bitte auch, welche Ansicht sie darstellt.

Zeichne bitte erst dann, wenn Du Dir über jeden Strich, den Du machen willst, in der Vorstellung bereits sicher bist.

A.5.5 Session 4

First block

T4.1 is similar to T3.2 except that it does not include the announcement of the follow-up test (T3.3)

Second block

T4.2:

Bitte bearbeite die Aufgabe solange, bis du selbst meinst, sie gut genug zu beherrschen, um einen dazugehörigen Test bestehen zu können. Wenn es so weit ist, kannst du selbstständig zum Test fortfahren. Solltest du allerdings feststellen, dass du den Test nicht ohne weiteres durchführen kannst, sondern noch weiter vorstellen musst, bitten wir Dich dies zu tun und deine Weiterarbeit daran wie zuvor zu protokollieren. Den Test solltest du also erst bearbeiten, wenn du dafür keine nennenswerte Denkarbeit mehr leisten musst.

Die Aufgabe:

*Falte ein Blatt in deiner Vorstellung **fünfmal** nacheinander, sodass es durch jede Faltung mittig halbiert wird, und die entstehenden Faltkanten senkrecht zueinander stehen.*

Mach dich mit dem Gegenstand nach jeder Faltung vertraut.

T4.3:

*Bitte stelle in je einer schematischen Skizze das **fünffach** gefaltete Blatt von links, rechts, oben, unten, hinten und vorne dar (insgesamt also 6 2D-Skizzen). Markiere unter der jeweiligen Skizze bitte auch, welche Ansicht sie darstellt.*

Zeichne bitte erst dann, wenn Du Dir über jeden Strich, den Du machen willst, in der Vorstellung bereits sicher bist.

APPENDIX B: STUDY GUIDELINES

The following guidelines should be followed by the investigator and are heavily based on our experience from the pilot studies. In their current form, they were used to guide the last two pilot cases (S5 and S6).

B.1 TRAINING SESSION

B.1.1 *Welcome & explanation of the session*

The subject is greeted in a friendly manner and the course of the session (as well as the next few days) is explained to them. Special care should be taken that they feel at ease with the experimental situation and that they are aware of their legal rights (namely that they may leave at any time, that any data collected will be only be published in anonymised or pseudonymised form, used solely for the purposes of scientific research, and that they have the right to withdraw their consent at any point in the future).

B.1.2 *Consent form*

The subject is handed a written form of consent, giving them a detailed account of their legal rights (cf. Appendix A). The investigator should inquire whether the subject understood everything on the form.

B.1.3 *General instructions*

The subject is shown the study software which will be used for all tasks except the first block in the training session (Figure 48). The first slide displayed on this browser-based interface contains the general instructions. The investigator should ask whether any of the instructions were not understood by the subject and explain them, if necessary.

B.1.4 *First task block*

For the first task block, the subject is informed that

1. in this block they will be presented *three short tasks* orally, consisting of consecutive instructions to imagine something
2. that they should attend to their progression of thoughts as they occur, and that
3. there will be a short, conversational interview afterwards, where we will try to “go back” to what had happened in their minds.



Figure 48: The general instructions as presented in the study software (“Allgemeine Hinweise”)

The first of the three tasks is given just to check whether the subject understands what they are supposed to do, i.e. without a subsequent interview.

After the second task, the first interview is conducted – based on Petitmen- gin (2006)’s concept of the *elicitation interview* (the interview for the third task should be conducted similarly):

- The interview starts with asking the subject to report what happened after the first instruction was given, using *direct quotes* of the instructions as recall stimuli), e.g.: “What happened after I gave you the instruction ‘Imagine the letter B’.”, “Next, I asked you to ‘Rotate (...)’.”, “What happened thereafter?”, etc.
- The subject can be encouraged to keep talking by saying things like “Go on” or “Please elaborate”, however care should be taken not to push the subject into reporting when they cannot remember or making them uncomfortable.
- If the investigator suspects that something the subject says might be speculative or some other post-hoc confabulation they might ask (in a neutral tone of voice) on which basis they have made this assertion (ideally quoting the subject).
- Generally, the main aim of the interview is to keep the subject talking with as little intervention as possible (and as much as necessary).
- If the subject uses vague language, in the form of “It was like X”, the investigator may ask “Can you elaborate on what you mean when you say X?” (with “No, I do not remember in more detail” being a valid response).
- If the subject keeps elaborating poetically on the contents of their mental images, the investigator should remind them that the focus of the report should be on the process, i.e. on what happened after another.
- The aim however should remain that the subject keeps talking and the investigator only “helps along” unfolding a few memories about the task.
- At the end of the interview the subject should arrive at a chronological report of what happened in their mind from the hearing the first instruction to recognising the resulting figure at the end.

- It might happen that the first of these reports will be rather vague and incomplete because the subject did not understand the instructions sufficiently or they did not pay enough attention introspectively. For this reason there is a third, similar task (and a second interview) where the subject might try once more.

B.1.5 *Debriefing for first block*

After the last interview about first block of tasks, there is a debriefing about the subject's general experiences with introspecting, imagery and reporting. More precisely the investigator should inquire after:

- Whether the subject noted something important about which we did not talk in the elicitation interviews
- Whether they felt at home with observing themselves or whether they were surprised by how it went
- How certain they felt about the assertions they made
- Referring to the "pitfalls" of introspections mentioned in the general instruction sheet, the investigator asks:
 - Whether they felt there were details that slipped past them (omission)
 - Whether they sometimes used words which immediately felt like they did not really fit (commission / inadequate descriptions)
 - Whether they found some things difficult to describe verbally and felt they could have expressed them otherwise more easily (overshadowing)
 - Whether they spend time speculating about the causes of events (speculative reasoning)

The aim of this debriefing is to make the subject feel at home with observing their thoughts and reporting on them, as well as familiarising them with the instructions pertaining to introspection. At this point (judging from experiences in the pilots), it is to be expected that the subject will have run into several of the "pitfalls" mentioned in the instructions. We assume that letting them experience these difficulties by themselves and thematising them in this interview (as opposed to informing them theoretically) will decrease the chance of those difficulties occurring later on in the main sessions.

After this interview, there is a short break of roughly five minutes. The subject should be reminded about the agenda after the break.

B.1.6 *Second task block*

The investigator tells the subject that

1. Now there is going to come a slightly longer, more complicated imagery task
2. That for this task, they are asked to fashion a *written* protocol (as for all subsequent tasks)
3. The main aim of working on this task is to familiarise them with this protocol technique

Crucially, the investigator tells the subject that in this setup, the investigator is sitting in front of another computer, *reading along* what they write in their protocol in real-time. The investigator explains that this is done to enable them to conduct the debriefing interview whose main aim it is to clarify passages that the investigator does not understand.

It is critical that the subject does not feel uncomfortable with this being the case. The investigator should mention that in this training session the subject might try out whether they would feel uneasy with this setup, and offer that an *alternative procedure* could be employed in the future sessions if this was the case¹.

The subject is asked to sit in front of the computer. The software then displays the protocol instructions (cf. Section 2.7) and prompts the subject to ask the investigator if they need more information or to click 'continue' otherwise (Figure 49).



Figure 49: The protocol instructions (“Hinweise zur selbstständigen Protokollerstellung”)

While the subject works on the task and fashions their protocol, the investigator should read along, and make written notes for their interview. They should note every occasion at which the protocol is unclear, ambiguous (with regard to the content of the report as well as the order of events), or where the investigator doubts the reliability of the report (e.g. speculation). They should also note other possibly relevant events like disturbances (e.g. noise), notable behaviour of the subject etc. However, they should take care not to distract the subject by observing them too intimately or by making an exuberant amount of notes.

¹ Namely: The investigator skims the protocol after it was fashioned, while the subject gets a 10 minute break. Note that pilots indicate that this procedure makes the debriefing interview far less valuable, as the investigator cannot identify every unclarity when reading an unknown protocol in such a short time

B.1.7 *Debriefing about second block*

The debriefing interview is based on the notes the investigator fashioned during the subject's work on the task. The aim of this interview is two-fold:

First, as with the preceding interviews, a chronological report of what happened during the task is sought after. The subject should be encouraged to talk and questions should remain of a clarifying nature and never be suggestive (always ask "What do you mean when you say X?" instead of "Do you mean A or B when you say X?").

Additionally however, the aim of this interview is to clarify *ambiguous or unclear passages* in the protocol. In order to do so, as the subject recounts what they observed chronologically, the investigator may "jump in" with an inquiry as the subject reaches the relevant passage. As with the elicitation interviews, the investigator should *quote the subject* and then ask whether they can provide a clarification. Unclearities can not only consist of misnomers, slips or vague descriptions, they can also pertain to an unclear chronology of events – however, judging from experience with the pilot subjects, reconstructing the correct sequence of events seems to be unsuccessful far more often than correcting other ambiguities.

B.1.8 *Final conversation*

Similar to the debriefing of the first task block, where the pitfalls mentioned in the instructions were thematised, in this part the investigator talks to the subject about the protocol they just fashioned. Going beyond these pitfalls, it is to be expected that the subject used a form of writing their protocols which does not lend itself particularly to protocol analysis. However, it would be far too much preliminary information for the subject to remember if all these points were mentioned in the beginning. Thus, we chose to let them write as it seems sensible to them and to only intervene *afterwards*, if their style would seriously hinder analysis. In particular, the following issues should be raised if they hold true for the protocol:

- *Staccato-like notes*: If the majority of protocol entries is only seconds apart, it is likely the subject mixed working on the task and reporting their observations. They should be reminded to separate these activities and not to rest their hands on the keyboard.
- *Very few, overly long notes*: Some subjects report only rarely and "result-oriented" instead of reporting their thoughts as they progress. They should be reminded that we are interested in their thoughts as they unfold over time and asked to be more attentive with respect to pauses they could take for reporting.
- *Very few notes overall*: Some subjects misunderstand the instruction to event-contingently retrospect to mean retrospectively very rarely and at the end of the main task. These protocols then only consist of a few sentences altogether. They should be explained that we are interested in them process-introspectively and asked to pause to report more often.
- *Overly literate language*: Some subjects confuse clarity of expression with eloquence and very arduously compose long-wound complex sentences (especially in German). They should be reminded that they might express themselves more colloquially, as long as they try to remain precise.

- *Overly abbreviating language*: On the other hand, some subjects take the instruction to remain short and precise too seriously and write only telegram-style notes. In itself this constitutes no problem, but sometimes these telegram-notes lend themselves to errors of commission (in trying to be as short as possible subjects use abbreviations and group together lots of possibly distinguishable phenomena).
- *Relative notation*: While working on a task, subjects themselves tend to know what they are attending to. In protocol analysis however, the analyser has to reconstruct what that was. If the protocol contains few direct denominations and instead a lot of pronouns, subjects should be reminded to explicitly name in the protocol what they are concerned with.
- *Other notational problems*: Furthermore, it might happen that subjects fashioned sketches but failed to name or number them and use these notations in their protocol.
- *Incomprehensible jumps*: Sometimes, it is not possible to reconstruct the course of events that led to something mentioned in the protocol because important preceding events are missing. The subject should be reminded that while can happen a lot and is still preferable over commission ("omission is better than commission"), they should if possible *make a note* in the protocol when something was omitted (either because they cannot remember properly or because they did not want to report a detail, e.g. for reasons of privacy).
- *Excessive reliance on sketching*: Especially with mental imagery tasks, some subjects, especially if they are graphically talented, might switch into a "drawing mode", where they express most of their thought progresses in sketches. They should be reminded that the primary means for reporting is the protocol, and sketches should only complement the protocol if need be.

B.2 MAIN SESSIONS

Since the structure of the training session intentionally mirrors that of the main sessions very closely, most of the guidelines given for the former equally hold for the latter, i.e. the debriefing interviews, presentation of instructions and tasks should be conducted similarly to the guidelines above. In the following, we will elaborate on what the investigator should take care of additionally, in the main sessions.

B.2.1 *Welcome*

Additionally to be welcomed in a friendly manner, from the second session onwards, the investigator should inquire from the subject whether they have been thinking about the paper-folding task *in the time between the sessions*. If this is the case, and the subject still remembers something about these thoughts, the investigator should conduct a short interview about them and record it.

B.2.2 *Instructions*

Before the first task block, the subject is shown the general and protocol instructions. As always, the investigator should ask the subject whether they

understood all the instructions. In later sessions, the subject should be reminded that even though they might feel very familiar with the instructions by now, they should nevertheless read them attentively. The investigator should suggest to the subject, that this might help them to get into the “introspective attitude” the instructions ask them to assume.

Before the second task block, the subject is shown the short reminder-version of the instructions, which have not been used in the training session.

B.2.3 *Selection of tasks*

Normally, the tasks presented to the subject should follow the overview in Section 2.5.3. However, if in the preceding session, the subject did not manage to fashion any sketches of the fold, or has shown major errors in their understanding of the fold, the number of folds should not be increased, and instead the preceding session should be repeated. Errors should be considered major, if to the investigator’s judgement, the subject would very likely not manage to imagine the more complex fold, and their understanding would benefit from a repetition of the easier task.

To avoid confusion on the subject’s side, the investigator should inform them, that today they will be given additional time to think about yesterday’s task.

B.2.4 *Debriefing interview*

The debriefing interview should be conducted in a similar manner than in the second block of the training session. However, since all main session tasks are from the same task domain, the investigator must pay particular attention to the way they speak about paper-folding. Generally, they should only talk about concepts and distinctions the subject has already talked about in their protocol or in the ongoing interview. Ideally, the investigator *uses the subject’s own vocabulary* for these concepts.

B.2.5 *Closing the session*

At the end of the session, the subject should be asked to try not thinking about or in any other way work on the task at home. The investigator should explain that even so, it might happen that during mind-wandering, they will think about the task. They should ask the subject to remember when this should happen and report about it at the beginning of the next session.

B.2.6 *Final remarks*

This concludes our guideline of how the investigator should conduct the study. Given that our task is deliberately ill-defined, and that each implementation of our study design is an individual case study, not all parts of the procedure can be as rigidly formulated as for a psychological experiment. In the end, the investigator has no choice but to use their own informed judgement, when reacting to a subject’s individual behaviour, and

deciding whether an error on the subject's side warrants repeating a session, or when conducting the debriefing interviews. To some extent, this contingency makes our study more like a Piagetian clinical interview or a Socratic dialogue than a strict psychological experiment (cf. Ginsburg, 1981; Stenning & van Lambalgen, 2004).

APPENDIX C: SOFTWARE

C.1 STUDY SOFTWARE

In this section we describe the browser-based study software which was developed with Moritz Jacobs. The software depends on current versions (as of mid-2015) of the following software packages: PHP, node.js, sass, and an IRC daemon, e.g. ngircd.

It has been tested on similarly recent versions of Mozilla Firefox, and might not display correctly on Google Chrome or other WebKit/Blink-based browsers. To display all the instructions and tasks correctly a minimal screen resolution of 1366×768 is required.

While there are plans to make the package publicly available, it needs further documentation and code clean-up for publication to be sensible. It is available at request from the thesis' author.

Any specific further dependencies of the software package can be handled by the npm manager, if node.js is installed correctly. After those have been installed, the software can be started with the included shell script `start.sh`. The browser interface will then be available under `http://127.0.0.1:8080`. After having been started, the software will ask the investigator to input a *subject number*, which will determine the IRC user name and hence the name of the log file, afterwards the software is ready to be used by the subject.

The slides it will display are read from the `slides/`-subdirectory and displayed in alphabetical order. Slide files are provided in *Markdown* format, allowing the use of rudimentary text layout options. The respective slide files for each session have to be placed manually in the `slides/`-subdirectory before starting the software.

Additionally, the software allows the following options:

If the first line of a slide file contains "`<btn:X>`", the text on the button changing to the next slide (which says "Weiter" per default) will change to *X*, allowing the use of more semantic button texts like "Yes, I have understood" or "To the next task". Since the subject should only advance the slides once they have read the current one carefully / finished with the task of the current one, there is no button allowing to change back to the previous slide. Should the next button have been pressed accidentally, the investigator can change back the the previous slide with the key combination "`Alt+.`".

Further options can be triggered by included certain keywords in the slides *file names*, namely:

`.cols2`: The text of the slide will be displayed in a two-column layout, which is useful for long instruction slides, like the ones displaying the general and protocol instructions.

.irc: The screen will be halved vertically, and the text of the slide will be displayed on the right half, while the left half displays the IRC text input field where the subject can type in their reports (used for the task slides).

The log files which contain the time-stamped reports the subject typed in and log messages of when they changed between slides (and hence tasks) are collected in in the subdirectory htdocs/logs/.

C.2 FOLDING IMPLEMENTATIONS

Recursive solution

The following code is a Prolog implementation of the recursive algorithm for generating Barbra representations of a fold's NV and DV, described in Algorithm 1. It was tested on a system with SWI-Prolog 7 (but it does not use any of the new features of version 7).

The main predicate is `recursivefold/4` which recursively calls `new_nv/3` and `new_dv/3` to construct the Barbra strings. Additionally it uses `vs/2` to generate the nested parentheses.

```
% base case: (for out- & in-fold) F1 is "||" and "|;"
recursivefold(_,1,"||","|;").

% recursive out-fold
recursivefold(out,Num,NV,DV) :- % exclude base case
    Num >= 2,
    DecrNum is Num - 1,
    % recursive call with decreased Num
    recursivefold(out,DecrNum,DecrNV,DecrDV),
    % build new out-folded NV from old DV
    new_nv(out,DecrDV,NV),
    % build new out-folded DV from old NV
    new_dv(out,DecrNV,DV).

% recursive in-fold
recursivefold(in,Num,NV,DV) :-
    Num >= 2,
    DecrNum is Num - 1,
    recursivefold(in,DecrNum,DecrNV,DecrDV),
    new_nv(in,DecrDV,NV),
    new_dv(in,DecrNV,DV).

% new_nv/3 generates NV of a fold based on the DV of the preceding fold
new_nv(_,";|","||||").

new_nv(out,OldDV,NewNV) :-
    % split old DV at the semicolon
    split_string(OldDV,";", "", [LeftOldDV,RightOldDV]),
    % get length of the right side of the old DV
    string_length(RightOldDV,LenRight),
    % integer-divide this length
    NumVs is LenRight // 2,
    % generate as many nested Vs (i.e. parens)
    vs(NumVs,NewVs),
    % concatenate: left-old-dv, right-old-dv, nested-vs, final single, closed V
    string_concat(LeftOldDV,RightOldDV,Part1),
```

```

string_concat(Part1,NewVs,Part2),
string_concat(Part2,"()",NewNV).

new_nv(in,OldDV,NewNV) :-
    % split old DV at the semicolon
    split_string(OldDV,";", "", [LeftOldDV,RightOldDV]),
    % get length of the left side of the old DV
    string_length(LeftOldDV,LenLeft),
    % integer-divide this length
    NumVs is LenLeft // 2,
    % generate as many nested Vs
    vs(NumVs,NewVs),
    % concatenate: left-old-dv, right-old-dv, nested-vs, final single, closed V
    string_concat(LeftOldDV,RightOldDV,Part1),
    string_concat(Part1,"()",Part2),
    string_concat(Part2,NewVs,NewNV).

% new_dv/3 generates DV of a fold based on the NV of the preceding fold
new_dv(out,OldNV,NewDV) :-
    % the new DV is the preceding NV, prepended by "||;" (a single, open V)
    string_concat("||;",OldNV,NewDV).

new_dv(in,OldNV,NewDV) :-
    % the new DV is the preceding NV, appended by "||" (a single, open V)
    string_concat(OldNV,"||",NewDV).

% Generate #Num numbers of nested parens
vs(0,"").
vs(Num,Vs) :-
    vs_helper(Num,V1List,V2List),
    atomic_list_concat(V1List,V1s),
    atomic_list_concat(V2List,V2s),
    string_concat(V1s,V2s,Vs).

% helper for vs/2, accumulating left- and right-parens tail-recursively
vs_helper(0,[],[]).
vs_helper(Num,["|Lparens"],["|Rparens"]):-
    NewNum is Num - 1,
    vs_helper(NewNum,Lparens,Rparens).

```

Iterative solution

The following code is a Prolog implementation of the iterative ("flat") algorithm for generating Barbra representations of a fold's NV (DV implementation not included here, but we could use `new_dv/3` from above), described in Algorithm 2. It was tested on a system with SWI-Prolog 7 (but it does not use any of the new features of version 7). The main predicate is `flatfold/3` which calls `even_foldseq/3` or `odd_foldseq/3`, and `incompose/3` or `outcompose/2` to construct the Barbra strings.

Even though the algorithm is much cheaper computationally, its even/odd case differentiations and reliance on numerical computation makes it a bit awkward to implement in Prolog. Additionally to the code given here, `vs/2` from above is needed to run properly.

```

flatfold(in,Num,NV) :- inflat(Num,NV).
flatfold(out,Num,NV) :- outflat(Num,NV).

```

```

% in-fold
inflat(1,"||").
inflat(Num,NV) :-
    % in case Num is odd
    Num mod 2 == 1,
    K is Num - 2,
    % call odd_foldseq for K=Num-2
    odd_foldseq(K,Seq),
    % call incompose/2 with this sequence
    incompose(Seq,NVlist),
    % concatenate the returned list and
    atomic_list_concat(NVlist,NVatom),
    % convert the result to a string
    atom_string(NVatom,NV).

% even case
inflat(Num,NV) :-
    Num mod 2 == 0,
    K is Num - 2,
    even_foldseq(K,Seq),
    incompose(Seq,NVlist),
    atomic_list_concat(NVlist,NVatom),
    atom_string(NVatom,NValmost),
    string_concat("||",NValmost,NV).

% outfold, odd case
outflat(1,"||").
outflat(Num,NV) :-
    Num mod 2 == 1,
    K is Num - 2,
    odd_foldseq(K,Seq),
    outcompose(Seq,NVlist),
    atomic_list_concat(NVlist,NVatom),
    atom_string(NVatom,NV).

% even case
outflat(Num,NV) :-
    Num mod 2 == 0,
    K is Num - 2,
    even_foldseq(K,Seq),
    outcompose(Seq,NVlist),
    atomic_list_concat(NVlist,NVatom),
    atom_string(NVatom,NValmost),
    string_concat("||",NValmost,NV).

% even_ and odd_foldseq/2 generate lists of even- or odd-numbered elements
% of OEIS sequence A027383, ranging from 0 to K
even_foldseq(K,X) :-
    % beware fencepost errors! range mod 2 == 1 yields 1,3,5,... seq_elems start at 0,
    % hence Ks mod 2 == 1 is (counter-intuitively) correct for even_foldseq/2
    findall(Ks,(between(0,K,Ks),Ks mod 2 == 1),ListofKs),
    maplist(seq_elem,ListofKs,X).

odd_foldseq(K,X) :-
    findall(Ks,(between(0,K,Ks),Ks mod 2 == 0),ListofKs),
    maplist(seq_elem,ListofKs,X).

```

```

% seq_elem/2 yields a single element of sequence A027383, helper for *_foldseq/2
seq_elem(0,1).
seq_elem(K,Result) :-
    NewK is K - 1,
    seq_elem(NewK,Temp),
    KthElem is 2^(floor(K/2)),
    plus(KthElem,Temp,Result).

% outcompose/2 takes a list of numbers (produced by one of the foldseq predicates)
% and generates the NV accordingly
outcompose(List,Compound) :-
    % call outcompose2/3 with ["|"] as initial element of the accumulator
    outcompose2(List,["|"],Compound).

% outcompose2/3 is an accumulator-based helper which _actually_ generates the NV
% base case: if list of numbers is empty, unify result ("Compound") with accumulator
outcompose2([],Compound,Compound).

% general case:
outcompose2([Head|Tail],Acc,Compound) :-
    % generate as many nested Vs as the first number in the list says
    vs(Head,Vs),
    % put a "|" in front of the current accumulator
    append(["|"],Acc,TempAcc),
    % put the Vs generated above at the end of this
    append(TempAcc,[Vs],TempAcc2),
    % and add a final "("
    append(TempAcc2,["("),NewAcc),
    % recursively call outcompose2/3 with remaining sequence and new accumulator
    outcompose2(Tail,NewAcc,Compound).

% same for in-folds:
incompose(List,Compound) :- incompose2(List,["|"],Compound).

incompose2([],Compound,Compound).
incompose2([Head|Tail],Acc,Compound) :-
    vs(Head,Vs),
    append(Acc,["|()"],TempAcc),
    append(TempAcc,[Vs],NewAcc),
    incompose2(Tail,NewAcc,Compound).

```

D

APPENDIX D: ABSTRACT

D.1 ABSTRACT

This thesis is an investigation of the question how a problem solver constructs the cognitive representations they use, and how they change over time. To this end, we first present an overview of the current state of problem solving theory, centred around Newell & Simon's classical problem space theory. In this overview, we note that research on the topic and development of the theory mostly concentrates on well-defined problems, identifying a lack of research on phenomena involving representational dynamics, such as learning and insight. The thesis subsequently claims that we need to investigate problems which are less straight-forward to comprehend (ill-defined problems). Such problems would in turn allow us the exploratory study of how subjects construct their cognitive representations of the problem.

In order to do so, we present a new study design together with a complex imagery manipulation and problem solving task (iterated, mental paper folding), and first observations from pilot studies.

The two main chapters of the thesis are (a) a description of the study design, methodology and the rationale behind it and (b) an analysis of the paper-folding task based on both, perfunctory observations from pilot studies, and formal investigations of the paper-folding domain. The task analysis chapter ends with a discussion of how our pilot studies' data could be described in terms of problem space theory.

Based on this assessment, we close the thesis with a discussion of potential directions for theory formation, lessons learned from the pilot studies, and an ideas regarding future research.

D.2 ZUSAMMENFASSUNG

Die vorliegende Arbeit behandelt die Frage wie ein Problemlöser die kognitiven Repräsentationen, die er/sie dafür einsetzt, konstruiert, und wie diese sich über die Zeit verändern. Zu diesem Zweck beginnen wir mit einem Überblick über den aktuellen Stand der Problemlöseforschung, mit primärem Fokus auf Newell & Simons klassische Problemraumtheorie. In diesem Überblick stellen wir fest, dass die Problemlöseforschung und zugehörige Theorieentwicklung sich stark auf sogenannte wohldefinierte Probleme beschränkt, und wir attestieren einen Mangel an Forschung zu Phänomenen, die repräsentationale Dynamiken, wie beispielsweise Lernen und Einsicht, involvieren. Die Arbeit fährt fort zu behaupten, dass es geboten wäre, Probleme zu beforschen, die weniger geradlinig aufzufassen sind („unscharf definierte Probleme“). Derartige Probleme würden es infolge gestatten, explorative Forschung darüber zu betreiben, wie Versuchspersonen ihre kognitiven Repräsentationen des Problems sukzessive konstruieren.

Zu diesem Zweck stellt die Arbeit einen neuartigen Versuchsaufbau und eine komplexe Vorstellungsmanipulations- und Problemlöseaufgabe vor (iteriertes, mentales Papierfalten), und präsentiert erste Beobachtungen aus Pilotstudien.

Die zwei Hauptkapitel der Arbeit sind (a) eine Beschreibung von Versuchsaufbau, Methodologie und dahinterstehender Grundgedanken, sowie (b) eine Aufgabenanalyse der Papierfaltaufgabe basierend sowohl auf vorläufigen Beobachtungen aus Pilotversuchen, als auch formalen Betrachtungen der Papierfalt-Domäne. Das Aufgabenanalyse-Kapitel schließt mit einer Diskussion darüber, inwiefern die Daten aus den Pilotversuchen im Sinne der Newell & Simonschen Problemraumtheorie verstanden werden können.

Auf dieser Betrachtung basierend schließen wir die Arbeit mit einer Diskussion über mögliche Entwicklungsrichtungen für Theorien menschlichen Problemlösens, Einsichten, die durch die Pilotstudien gewonnen wurden, und Ideen für potentielle weiterführende Forschung.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both $\text{\LaTeX}$ and $\text{\LyX}$:

<https://bitbucket.org/amiede/classicthesis/>