UNIVERZA V LJUBLJANI
SKUPNI INTERDISCIPLINARNI ŠTUDIJSKI PROGRAM DRUGE STOPNJE
KOGNITIVNA ZNANOST
V SODELOVANJU Z UNIVERSITÄT WIEN, UNIVERZITA KOMENSKÉHO V
BRATISLAVE IN EÖTVÖS LORÁNDTUDOMÁNYEGYETEM
IN INŠTITUTOM ZA EKSPERIMENTALNO PSIHOLOGIJO
SLOVAŠKE AKADEMIJE ZNANOSTI

NEJC GRENC

Vpliv Mentalnih Reprezentacij na Postopek Reševanja Problemov

MAGISTRSKO DELO

11 March 2021

UNIVERSITY OF LJUBLJANA
JOINT INTERDISCIPLINARY MASTER'S STUDY PROGRAM
COGNITIVE SCIENCE
IN COOPERATION WITH UNIVERSITÄT WIEN, UNIVERZITA KOMENSKÉHO V
BRATISLAVE IN EÖTVÖS LORÁNDTUDOMÁNYEGYETEM
AND INSTITUTE OF EXPERIMENTAL PSYCHOLOGY
OF SLOVAK ACADEMY OF SCIENCES

NEJC GRENC

Vpliv Mentalnih Reprezentacij na Postopek Reševanja Problemov

The Role of Mental Representation in Problem-Solving Process

MASTER'S THESIS

External mentor:
dr. Matúš Grežo
Researcher at Institute of
Experimental Psychology
Slovak Academy of Sciences,
Bratislava, Slovakia

Home mentor:
prof. dr. Grega Repovš
Associate Professor in
Cognitive Psychology,
Faculty of Arts,
Ljubljana, Slovenia

11 March 2021

Abstract

Although the mental representation of a problem is an often-overlooked aspect in the problem-solving research, it has a significant effect on the understanding of the problem, on strategy selection, and on problem-solving performance. We have performed an experimentally-driven online study with help from 254 participants, which focused on one of the aspects of mental representation research, namely the transfer of mental representation during a problem-solving process. We achieved this by using matchstick tasks, where problem-solvers had to correct the matchstick equations by moving the least number of matchsticks as fast as possible. Before the experiment, problem-solvers learned and trained one of the possible problem-solving strategies, leading to the creation of the associated mental representation. In the testing process, they were provided with a sequence of tasks, where each subsequent task was less optimally solvable by the learned strategy than the previous one, slowly forcing the problem-solvers into an impasse, where they needed to change their strategy in order to continue solving problems. This way, by manipulating tasks with different associated mental representations, we investigate how problem-solvers' mental representations, created in previous tasks, induce positive or negative transfer and thus in subsequent tasks affect their problem-solving performance in terms of time and the number of moves taken per task. Based on the results of this study, we conclude that mental representations differ significantly in their availability, with the representation of moving matchsticks between or within numerals having higher availability than the representation of moving matchsticks between or within operators. Transitively, their inherent differences exhibit a significant, positively-correlated effect on transfer and performance in problem-solving processes, unexpectedly even in control groups. While problems themselves are independent from problem-solvers, their solutions and the process of achieving them are not. In this thesis we show that this often-overlooked aspect of problem-solver's initial mental representation has a significant role on the problem-solving process.

Keywords

- problem-solving
- mental representation
- matchstick equation
- online experiment

Povzetek

Čeprav je mentalna reprezentacija problema pogosto spregledan vidik v raziskavah reševanja problemov, ima pomemben vpliv na razumevanje, izbiro strategij in uspešnost reševanja problema. Izvedli smo eksperimentalno spletno raziskavo s pomočjo 254 udeležencev, ki se je osredotočila na enega od vidikov raziskav mentalnih reprezentacij, in sicer na prenos (ang. transfer) mentalne reprezentacije med postopkom reševanja problemov. To smo dosegli z uporabo vžigaličnih enačb, kjer so morali udeleženci čim hitreje popraviti nepravilne vžigalične enačbe s premikom najmanjšega števila vžigalic. Pred poskusom so se udeleženci naučili in trenirali eno od možnih strategij za reševanje problemov, kar je vodilo k ustvarjanju povezane mentalne reprezentacije. Med procesom testiranja jim je bilo dano zaporedje vžigaličnih nalog, kjer je vsaka naslednja naloga manj optimalno rešljiva z naučeno strategijo kot prejšnja. Tako so bili reševalci problemov počasi prisiljeni v mentalni zastoj (ang. impasse), kjer so morali spremeniti svojo strategijo, da so lahko nadaljevali z eksperimentom. Tako z manipulacijo nalog z različnimi povezanimi mentalnimi reprezentacijami raziskujemo, kako mentalne reprezentacije reševalcev problemov, pridobljene z reševanjem prejšnjih nalog, sprožijo pozitiven ali negativen prenos in tako vplivajo na uspešnost reševanja nadaljnjih nalog v času in številu potez, ki jih reševalec potrebuje za rešitev naloge. Naši rezultati nakazujejo, da se mentalne reprezentacije bistveno razlikujejo po svoji razpoložljivosti, pri čemer je razpoložljivost premikov vžigalic med številkami ali znotraj njih večja kot razpoložljivost premikov vžigalic med operatorji ali znotraj njih. Tranzitivno pa razlike med njimi kažejo pomemben, pozitivno koreliran učinek na prenos in uspešnost reševanja problemov, nepričakovano tudi v kontrolnih skupinah. Čeprav so problemi neodvisni od reševalcev problemov, njihove rešitve in postopek njihovega uresničevanja niso. V tem delu pokažemo, da ima ta pogosto spregledan vidik začetne mentalne reprezentacije reševalca problema pomembno vlogo pri postopku reševanja problemov.

Ključne besede

- reševanje problemov
- mentalna reprezentacija
- vžigalična enačba
- spletni eksperiment

Declaration of Authorship

I, Nejc GRENC, declare that this thesis titled, "The Role of Mental Representation in Problem-Solving Process" and the work presented in it are my own. I confirm that:

- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- This work was done for the sole purpose of candidature for a master's degree at the University of Ljubljana.

Acknowledgements

The research presented in this thesis has been conducted in a small team of three people, consisting of the author of this thesis and the two mentors, in a slow pace over a span of a few years. I am very grateful for this experience, as I have learned a lot about scientific practices, work ethics and project management.

Firstly, I would like to say "Ďakujem mnohokrát" to my external mentor dr. Matúš Grežo from the Institute of Experimental Psychology of Slovak Academy of Sciences. At our first encounter, he gave me the opportunity to conduct my research, even after I had presented him the over-ambitious and ill-designed research plan. The research has developed and matured over time into a proper scientific study under his guidance. He continued to encourage my endeavors even through the long stretches of silence during the development cycles and his support never wavered through the duration of the research. I am very thankful for his patience and the support he has provided me over the years. On this note, I would also like to thank prof. ing. dr. Igor Farkaš from Centre for Cognitive Science at Comenius University in Bratislava for introducing me to my external mentor.

Prav tako se najlepše zahvaljujem tudi domačemu mentorju izr. prof. dr. Gregi Repovšu s Filozofske fakultete pod okriljem Univerze v Ljubljani, ki mi je med študijem ustvaril zanimanje za psihološke pojave, ki me je vodilo skozi celoten projekt. Zahvaljujem se mu za potrpežljivost in pomoč skozi celoten projekt.

I would like to thank dr. Paolo Petta from Austrian Research Institute for Artificial Intelligence for his support during a preceding study, in which he introduced me to good project management practices and documentation keeping. Thanks to his teachings, I never gave up on the project despite its long duration.

An immense "thank you" goes to Karina Rerichová for her unwavering support and help throughout the research. She shared my joy through enthusiastic times of launching the website and obtaining our first successful results, and has encouraged me through many tough times of despair during the stretched-out research process. Without her words of encouragement, I would probably have not completed this thesis. Additionally, I would like to thank her for the help in acquiring subjects for the study and for reviewing my thesis. Thank you, Karina!

In ne nazadnje se zahvaljujem svoji družini za nenehno spodbujanje skozi daljše obdobje tega projekta. Vaša podpora me je držala nad vodo vsa ta študentska leta, zato sem vam neizmerno hvaležen. Hvala!

Razširjeni povzetek

Kakor pravi stari pregovor: "vaja dela mojstra", trening izboljšuje reševalne sposobnosti in reševalcu omogoča, da hitreje in enostavneje reši podobne probleme. Vendar ni vedno tako. Prenos znanja pri reševanju problemov (angl. "transfer"), kot ga pogosto omenjajo v geštalt psihologiji, je učinek naučenega znanja na nov problem. Prenos normalno izboljša, v določenih situacijah pa lahko tudi ovira sposobnost reševanja novega problema (Luchins, 1942; Saugstad and Raaheim, 1960).

Eden od odločilnih vidikov prenosa je mentalna reprezentacija danega problema, ki je ustvarjena s kombinacijo opisa problema in preteklimi izkušnjami. Ta mentalna reprezentacija določa, kako reševalec problema razume problem in igra pomembno vlogo pri izbiri razpoložljivih strategij. Vendar pa posamezna mentalna reprezentacija morda ne privede do optimalne rešitve in lahko celo blokira proizvodnjo uspešne rešitve, saj je reševalci pogosto ne morejo spremeniti, da bi se lahko izognili negativnemu prenosu znanja. Mentalna reprezentacija pomembno vpliva na prenos ter tako oblikuje postopek reševanja problema (Kotovsky and Fallside, 1989).

Opredelitev problema

Vendar pa ni nujno, da reševalec med reševanjem problema ostaja zvest posamezni mentalni reprezentaciji. Mentalno reprezentacijo je mogoče zamenjati z drugo in pri tem spremeniti razumevanje problema in izbiro strategij ter tako oblikovati nov postopek reševanja problema. Ustvarjanje nove mentalne reprezentacije je zahteven mentalni proces (Wertheimer and Wertheimer, 1959). Ko je mentalna reprezentacija ustvarjena, jo reševalec zlahka uporabi. Reševalec lahko v mislih drži le eno mentalno reprezentacijo naenkrat, vendar jo lahko kadar koli zamenja z drugo že znano (že ustvarjeno) mentalno reprezentacijo.

Dober vizualen primer mentalnih reprezentacij oriše slika 1, na kateri je mogoče razpoznati dve podobi: mlado damo in starejšo gospo. Prvič, ko opazovalec vidi to sliko, si samodejno ustvari eno od podobi-pripadajočih mentalnih reprezentacij, medtem ko razpoznavanje druge podobe in ustvarjanje pripadajoče mentalne reprezentacije zahtevata aktivno razmišljanje. Ko pa opazovalec ustvari obe mentalni reprezentaciji, ju lahko zamenjuje skoraj brez truda.

SLIKA 1: Slika prikazuje dve mentalni reprezentaciji: mladenko in staro gospo.

(Author: Hill, W. E., Title: "My wife and my mother-in-law. They are both in this picture - find them", Published: Puck, v. 78, no. 2018 (1915 Nov. 6), p. 11. Image source: https://insolemexumbra.files.wordpress.com/2015/07 /young_lady_old_woman_illusion.jpg)

Mentalne reprezentacije se prenašajo med podobnimi problemi in z njimi vred tudi strategije reševanja. Ne velja pa vedno, da je strategija reševanja, ki je uspešna v reševanju določenega problema, uspešna tudi pri reševanju podobnih problemov. V primeru, da prenos mentalne reprezentacije med dvema problemoma omogoči uspešnejše reševanje drugega problema, govorimo o pozitivnem prenosu. Ko takšen prenos mentalne reprezentacije negativno vpliva na uspešnost ali celo na zmožnost reševanja drugega problema, pa govorimo o negativnem prenosu.

V naši študiji se osredotočamo na raziskovanje, kako prenos različnih mentalnih reprezentacij vpliva na uspešnost reševanja problemov. Globlje razumevanje povezave med prenosom in mentalnimi reprezentacijami lahko reševalce problemov vodi v izbiro optimalne mentalne reprezentacije, ki jim omogoči najuspešnejše reševanje določenih problemov.

Kratko teoretično ozadje

Reševanje problemov raziskuje precej znanstvenih področij s številnih različnih vidikov. Raziskovalci pogosto proučujejo probleme v smislu njihove zasnove, zapletenosti, preglednosti, dinamike in raznolikosti ciljev. Vse z namenom, da bi čim bolj razumeli probleme ter tako izboljšali strategije in tehnologije z namenom najti boljše in hitrejše rešitve. Raziskovanje in razumevanje problemov sta sama po sebi problem, ki se ga je treba lotiti. Eden od bolj zapostavljenih vidikov pa je vpliv mentalnih reprezentacij na sam proces reševanja problemov.

Mentalna reprezentacija problema je notranja podoba ali miselni model, kako problem zaznava reševalec problemov. Različne mentalne reprezentacije lahko na različne načine oblikujejo misli in dejanja reševalcev, čeprav je problem v vseh pogledih popolnoma enak. Mentalne reprezentacije raziskuje tudi Gestalt šola psihologije, kjer mentalno reprezentacijo imenujejo "geštalt". Izraz "geštalt" se nanaša na organizirano celoto elementov in njihove medsebojne povezave v strukturi.

Dodatno velja, da so mentalne reprezentacije neopisne zaznavne entitete, ki so specifične za vsakega posameznika. Niso prenosljive med posamezniki, zato jih je nemogoče v celoti opisati z besedami. Mentalne reprezentacije same je zelo težko znanstveno proučevati. Vendar to ne pomeni, da ni mogoče proučiti njihovih učinkov na psihološke vidike, kot je vedenje ali v našem primeru reševanje problemov. V tem pogledu so podobne barvam, ki so prav tako neopisne (modro barvo je nemogoče opisati osebi, ki modre barve še ni videla), obstaja pa veliko študij, ki potrjujejo učinke različnih barv na čustvena stanja (npr. Strapparava and Özbal, 2010).

Newell, Simon, et al. (1972) uvedejo teoretični okvir za opis reševanja problemov, v katerem reševalec problema iz nabora vseh možnih mentalnih reprezentacij ustvari najmanj eno reprezentacijo problema in z njo povezan problemski prostor, ki predstavlja vsa možna stanja problema. V tem problemskem prostoru reševalec izvaja usmerjene operacije (imenovane "premiki") in se z vsako operacijo premika skozi različna stanja problema od začetnega do ciljnega stanja (Kotovsky and Fallside, 1989).

Ker človeški um v nasprotju z računalniki deluje asociativno, ni optimiziran za pregled in iskanje celotnega problemskega prostora. Da se izognemo tej omejitvi, se pogosto zatekamo k uporabi strategij in/ali hevristik, ki lahko odločilno vplivajo na postopek in izid reševanja problema. Z uporabo strategije ali hevristike se reševalec osredotoči le na del že znanega problemskega prostora, kjer ve, kako najti rešitev.

Uspešnost postopka reševanja problemov je tako v posameznikovem dojemanju problemskih elementov in razpoložljivih premikih, ki so integrirani v posameznikovo mentalno reprezentacijo problema (Kotovsky and Fallside, 1989).

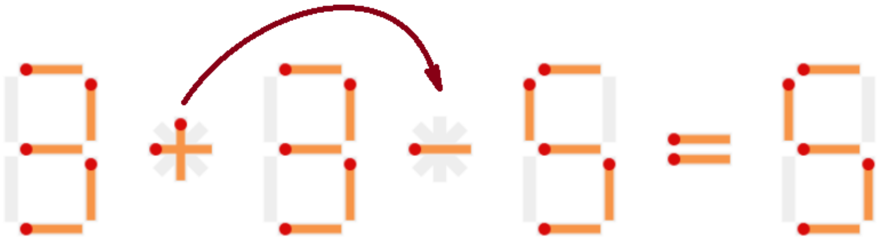
Kot že omenjeno, se mentalna reprezentacija ohranja med podobnimi primeri. Ko reševalec prepozna strategijo reševanja določenega problema, poskuša to strategijo uporabiti tudi na nadaljnjih podobnih problemih. V primeru, da je tudi naslednji podoben problem rešljiv s to strategijo, ta prenos mentalne reprezentacije pospeši reševanje problema, saj reševalcu prihrani čas ustvarjanja nove mentalne reprezentacije. Tak prenos imenujemo pozitiven prenos.

Nasprotno pa velja v primeru, da naslednji podoben problem ni rešljiv s strategijo obstoječe mentalne reprezentacije. V tem primeru reševalec vseeno (neuspešno) poskusi rešiti problem z obstoječo strategijo. Za uspešno reševanje pa mora spoznati, da obstoječa strategija ne deluje, premagati zastoj ter ustvariti novo mentalno reprezentacijo in strategijo, s katero bo dejansko lahko rešil ta problem. Tak prenos močno podaljša čas reševanja, zato ga imenujemo negativen prenos (Kotovsky and Fallside, 1989).

Velik del mentalnih reprezentacij predstavlja proces njihovega ustvarjanja ali spreminjanja, saj ima pomemben in merljiv učinek na uspešnost v procesu reševanja problemov (Kotovsky and Fallside, 1989). Ko oseba poskuša prepoznati ali zamenjati mentalno reprezentacijo, česar še nikoli ni storila, doživi zastoj (angl. "impasse"). Zastoj je mentalno stanje, ko so reševalci problemov "zataknjeni" in ne morejo napredovati v svojem trenutnem procesu reševanja problemov. Pogosto ga spremlja frustracija, ki jo je mogoče rešiti z vpogledom (angl. "insight") (Ohlsson, 1992). Premagovanje zastoja zahteva zavedno in dolgotrajno prizadevanje ter tako vodi do zapoznelega odziva, zato ga je tudi mogoče znanstveno proučevati. Ko reševalec premaga zastoj, postane mentalna reprezentacija razpoložljiva in reševalec lahko v nadaljnjih problemih olajša zamenjavo mentalne reprezentacije, ne da bi mu bilo treba ponovno premagati takšen zastoj.

Knoblich et al. (1999) so izvedli serijo eksperimentov, v katerih so morali udeleženci rešiti preproste probleme vžigaličnih enačb. Udeleženci so dobili matematično napačne (vendar veljavne) vžigalične enačbe z rimskimi številkami in preprostimi aritmetičnimi operacijami. Udeleženci so jih morali rešiti/popraviti s premikanjem ene vžigalice. Poleg tega njihovo delo dodatno podpira Alzayat (2011), ki je izvedel podoben eksperiment, v katerem so morali udeleženci rešiti vžigalične enačbe s premikom vžigalice med števkami ali s premikom vžigalice med operatorji. Te različne vrste premikov se imenujejo strategije in izbira posamezne strategije odločno vpliva na postopek reševanja teh problemov.

V naši študiji razširimo zasnovo eksperimenta vžigaličnih enačb, ki so ga ustvariliAlzayat (2011) in Knoblich et al. (1999), z novim sklopom nalog in drugačnimi spremenljivkami ter njihove ugotovitve povežemo z mentalnimi reprezentacijami (Kotovsky and Fallside, 1989). Slika 2 prikazuje primer reševanja ene izmed vžigaličnih enačb. V okviru našega vžigaličnega eksperimenta raziskujemo, ali mentalne reprezentacije vplivajo na uspešnost reševanja problemov in kako močan je ta učinek.



SLIKA 2: Primer reševanja vžigalične enačbe

Metodologija

V naši študiji smo raziskali značilnosti ustvarjanja in zamenjave mentalnih reprezentacij. Udeleženec je bil sprva naučen mentalne reprezentacije ali pa jo je moral ustvariti sam. Nadalje je reševal podobne naloge, kjer je bil postopoma prisiljen v zastoj ter tako primoran spremeniti svoje dojemanje problema in zamenjati mentalno reprezentacijo.

Glavni cilj naše študije je proučiti učinek (udeležencevih) naučenih strategij na hitrost in uporabo potez med postopkom reševanja serije vžigaličnih nalog in ugotoviti, v kolikšni meri posamezniki prenesejo že naučene strategije na naslednje naloge. Ta prenos je merljiv vidik mentalne reprezentacije, povezane s to strategijo.

V našem eksperimentu smo uvedli dve specifični strategiji:

> ***Strategija 'a'*** – premikanje vžigalice z ene števke na drugo števko ali znotraj ene števke;
>
> ***Strategija 'b'*** – premikanje vžigalice z enega operaterja na drugega ali znotraj enega operaterja.

Vsaka naloga v našem eksperimentu je specifično izbrana tako, da je rešljiva z eno ali z drugo strategijo. Vse izbrane vžigalične enačbe so na pogled videti podobno, a so rešljive na različne specifične načine. Podoben videz enačb omogoča prenos mentalnih reprezentacij, način njihovega reševanja pa določi, ali pride do pozitivnega ali negativnega prenosa.

Za namen merjenja prenosa smo izbrali 10 vžigaličnih enačb in jih razporedili v dve zaporedji. V prvem zaporedju so naloge razporejene začenši s tistimi, ki so rešljive izključno s strategijo 'a', nadalje s tistimi, ki so rešljive z obema strategijama enako optimalno, in nazadnje s tistimi, ki so rešljive izključno s strategijo 'b'. V drugem zaporedju so iste naloge razporejene v obratnem vrstnem redu.

Ti zaporedji sta specifično izbrani tako, da v prvem delu eksperimenta udeleženec rešuje naloge s strategijo, ki mu je že znana. Pri tem se ustvarjena mentalna reprezentacija prenaša med nalogami in tako spodbuja pozitiven prenos. V drugem delu eksperimenta pa ta strategija postopoma postaja vedno manj optimalna in nazadnje privede do zastoja, kjer nujno pride do zamenjave mentalne reprezentacije in negativnega prenosa. Udeležence smo naključno razporedili med obe zaporedji.

V naši študiji smo izmerili tudi vpliv predhodnega znanja. Vsakega izmed udeležencev smo bodisi dodelili eksperimentalnim skupinam in ga naučili začetno strategijo (strategijo, s katero je rešljiva prva naloga danega zaporedja) bodisi dodelili kontrolnim skupinam in ga naučili obe strategiji ali pa nobene od njih.

Na ta način smo ustvarili 6 skupin udeležencev (2 eksperimentalni in 4 kontrolne; po 3 za vsako zaporedje). S to kombinacijo smo izmerili vpliv ustvarjanja začetne reprezentacije (vpliv na uspešnost reševanja prve naloge) in vpliv zastoja (vpliv na uspešnost reševanja naloge, ko udeleženec zamenja mentalno reprezentacijo).

Ne nazadnje, naš eksperiment je bil izveden na spletu. Digitalna distribucija nam je omogočila širše občinstvo in s tem povečala pričakovano število udeležencev naše študije. Udeleženci so lahko sodelovali v študiji iz domačega udobja in več udeležencev je lahko sodelovalo hkrati. Poleg tega je bila vsaka eksperimentalna seja izvedena natančno po vnaprej določeni opredelitvi, s čimer smo odpravili vse možne vplive eksperimentatorjev in prihodnjim znanstvenikom in raziskovalcem omogočili, da proučijo zasnovo eksperimenta do najmanjših podrobnosti izvedbe. Eksperiment je še vedno razpoložljiv na e-naslovu `matchstick-task.eu`.

Rezultati

Za izvajanje naše študije smo ustvarili 3 hipoteze in raziskovalno vprašanje:

1) Posamezniki, ki se naučijo dveh strategij, naloge rešijo uspešneje v primerjavi z udeleženci, ki se naučijo samo ene ali nobene strategije.
2) Prenos je največji pri posameznikih, ki se naučijo samo ene strategije reševanja nalog.
3) Posamezniki, ki se naučijo vsaj ene strategije, prvo nalogo rešijo hitreje kot posamezniki, ki se ne naučijo nobene od strategij.

- Ali je ena strategija bolj razpoložljiva (pogosteje uporabljena) kot druga? Ali katera strategija vodi do večjega prenosa kot druga?

Te hipoteze proučijo vse pomembne vidike prenosa mentalnih strategij. Prva hipoteza proučuje vpliv zastoja. Druga hipoteza proučuje moč prenosa posamezne mentalne reprezentacije. Tretja hipoteza proučuje vpliv ustvarjanja začetne mentalne reprezentacije. Raziskovalno vprašanje pa proučuje razlike med izbranima mentalnima reprezentacijama.

Uspešnost reševanja je bila izmerjena s številom uporabljenih potez in porabljenim časom, ki ju je reševalec porabil za reševanje posamezne naloge. Večje število potez ali večja količina porabljenega časa pomeni manj uspešno reševanje naloge. V opisni analizi smo dodatno potrdili, da sta količina potez in časa pozitivno korelirana. Vsaka statistična analiza uspešnosti je bila dejansko izvedena posebej na številu uporabljenih potez in posebej na porabljenem času, a bomo zaradi zgoščenosti tega povzetka, ti dve vrednosti še naprej naslavljali kot "uspešnost".

Statistična analiza izmerjenih vrednosti je pokazala nenormalno porazdelitev. Zato smo morali uporabiti neparametrične statistične teste, ki so lahko analizirali nenavadno porazdeljene podatke. Za našo analizo rezultatov smo uporabili Mann-Whitney test, Wilcoxon Signed-Rank test ter ANOVA in MANOVA teste, ki so zmogli analizirati različne velikosti vzorcev s heteroscedastičnimi variancami. Za pridobitev pomembnih rezultatov ANOVA in MANOVA smo med skupinami izvedli tudi večvariatne parne post-hoc teste. S temi testi smo lahko v celoti analizirali vse podatke, potrebne za naše hipoteze.

Pri analizi podatkov za prvo hipotezo nismo zaznali, da bi skupini, ki sta bili v začetku eksperimenta naučeni obeh strategih, značilno uspešneje rešili katere koli izmed nalog.

Drugo hipotezo smo razdelili na tri delovne manjše hipoteze in posebej izmerili razmerje prenosa, pozitiven prenos in negativen prenos. Pri analizi razmerja prenosa se je izkazalo, da so kontrolne skupine doživele podoben prenos (brez značilnih razlik) kot eksperimentalni skupini. Pri analizi pozitivnega prenosa smo opazili, da so vse skupine doživele močan pozitiven prenos, saj se je uspešnost reševanja povečevala z vsakim naslednjim podobnim primerom (ki je rešljiv s strategijo prejšnjega) pri vseh skupinah. Nismo pa opazili, da bi kontrolne skupine, ki teoretično doživijo največjo količino pozitivnega prenosa, značilno uspešneje rešile te naloge.

Pri analizi negativnega prenosa smo za vsakega udeleženca v eksperimentalnih skupinah razpoznali "točko dekompozicije", kjer se je udeleženec spoprijel z zastojem in ustvaril novo mentalno reprezentacijo. Z analizo smo dokazali, da je udeleženec v točki dekompozicije rešil nalogo značilno manj uspešno kot udeleženci v kontrolnih skupinah, ter tako potrdili hipotezo, da premagovanje zastoja in ustvarjanje nove mentalne reprezentacije zmanjšata uspešnost reševanja naloge. Med analizo pa smo opazili tudi, da udeleženci z začetno strategijo 'b' to zamenjajo skoraj takoj, ko nalog ne rešijo več optimalno uspešno, medtem ko udeleženci z začetno strategijo 'a' to strategijo ohranjajo dlje, ne glede na to, da nalog ne zmorejo optimalno uspešno rešiti.

Pri analizi tretje hipoteze pa smo ponovno zaznali neznačilne razlike med skupinami. Vse skupine enako učinkovito rešijo svojo prvo nalogo, ne glede na to, da so morale nekatere skupine pri tem ustvariti začetno mentalno reprezentacijo. To nakazuje, da je ustvarjanje začetne mentalne reprezentacije nezahteven proces.

Pri analizi raziskovalnega vprašanja smo primerjali uporabo strategij 'a' in 'b', kjer se je izkazalo, da v splošnem udeleženci pogosteje uporabljajo strategijo 'a' tudi pri nalogah, kjer sta obe strategiji enako optimalno uspešni. Dodatno je večja razpoložljivost strategije 'a' podprta z rezultati analize negativnega prenosa, kjer smo pokazali, da udeleženci uporabljajo strategijo 'a' tudi v primerih, kjer ni več optimalno uspešna. Zaključimo, da je mentalna reprezentacija strategije 'a' bistveno bolj razpoložljiva kot mentalna reprezentacija strategije 'b' (premikanje vžigalice med števkami ali znotraj njih je bilo lažje/bolj razpoložljivo kot premikanje vžigalice med operaterji ali znotraj njih).

Sklep

Pridobljeni rezultati niso potrdili vseh naših hipotez. Namesto tega pa so pokazali, da so celo kontrolne skupine prikazale izrazit učinek prenosa. Prav tako v nasprotju z našimi pričakovanji daljše učenje mentalnih reprezentacij in strategij ni pokazalo izboljšanega prenosa začetne mentalne reprezentacije. Izkazalo se je, da **učenje, ponavljanje in trening ne pomenijo večjih izboljšav v uspešnosti, kot je bilo pričakovano**.

Nadalje smo opazili, da je razpoložljivost mentalne reprezentacije pozitivno povezana z negativnim prenosom in da udeleženci uporabljajo mentalno reprezentacijo z večjo razpoložljivostjo tudi pri nalogah, za katere je bila ta manj optimalna. Velja, da **se posamezniki držijo svojih obstoječih in pogosto uporabljenih vedenj in praks**.

Iz naših opažanj smo ugotovili, da se je uspešnost reševanja na mestu zamenjave mentalne reprezentacije znatno zmanjšala. To je bilo še posebej očitno pri bolj razpoložljivi mentalni reprezentaciji. Tako sklepamo, da **najbolj razpoložljive mentalne reprezentacije vodijo v najdražje mentalne zastoje**.

Naša študija je bila uspešna. Nismo dobili rezultatov, ki smo jih želeli, a smo dobili rezultate, ki nudijo dragocen vpogled v vlogo mentalnih reprezentacij pri reševanju problemov.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Every problem has a solution.
You just have to be creative enough to find it."
– Travis Kalanick

Being alive means experiencing challenges and problems on a daily basis. Such problems vary from as simple as selecting a meal, to more complicated, like getting married or choosing a place to live. In an effort to continue or even improve our lifestyle, we have to tackle and solve these problems. Problem-solving is a state of mind that occurs during the process of tackling the problem, when the mind is focused on finding an elusive solution.

A common proverb found in many cultures and languages, i.e. *"Practice makes perfect"* (*"Vaja dela mojstra"* (SI), *"Übung macht den Meister"* (DE), *"Opakovanie je matka múdrosti"* (SK), etc.) tells us that past experience can improve our problem-solving abilities. This proverb's interpretation that 'solving a problem is simpler for the second time' is also directly attributed to learning. Well-learned or practised problem-solvers often perform better at solving problems. Following this advice, to enhance our problem-solving skills, we should only practice and learn as much as possible. However, reality is much more complicated and, as we find out in this thesis, in certain cases the exact opposite is true. Sometimes past experience and knowledge actually makes us worse problem-solvers.

Effective experience or transfer in problem-solving, as it is often referred to in Gestalt psychology, is an effect of learned/accumulated knowledge on the new problem and it can turn out to be positive or negative; it can enhance our ability to solve future problems or hinder it respectively (Saugstad and Raaheim, 1960; Luchins, 1942). (Knowledge) Transfer marks how well problem-solvers are able to solve subsequent problems after solving the initial ones. Positive transfer results in problem-solvers solving future problems better because of having solved previous problems, while negative (unwanted) transfer indicates that problem-solvers perform worse, because previous experience is applied to problems that require different solutions or problem-solving strategies. The term 'transfer' is used in the problem-solving research field and can be (over-)simplified as a combination of solving a problem and applying the newly obtained knowledge to the next problem. While the effect of transfer has been deeply researched, the inner workings of transfer are still under research.

One of the determining aspects of transfer is the mental representation of a given problem, created by a combined problem description and past experience. This mental representation determines how problem-solvers perceive a problem and plays an important role in the selection of available strategies and moves. It subsequently shapes the problem-solving process and the obtained solution (Kotovsky and Fallside, 1989). However, the initial mental representation, which is created at the beginning of a problem-solving process, may not lead to an optimal result and can even block the production of a successful solution, as subjects are often unable to change it in order to avoid the negative knowledge transfer.

Following Simon's work (Kotovsky and Fallside, 1989), mental representations can have a significant effect on transfer. In our study, we expand on the experiment design by Alzayat (2011) and Knoblich et al. (1999), which uses matchstick-equation problems, with a new set of tasks and different variables, and connect their findings with mental representations (Kotovsky and Fallside, 1989). Within the scope of our matchstick experiment, we investigate whether mental representations have any effect on problem-solving performance and how strong this effect is. We measure this effect in terms of time and the number of moves taken per task between six different groups in which participants have learned between zero and up to two different mental representations. We are especially interested in the positive and negative transfer features of mental representations, therefore we manipulated the equations encountered by each group. The first couple of tasks are solvable optimally using the initially learned strategy and thus facilitate positive transfer. Such solutions become less optimal (requiring more time and moves to solve) for each subsequent task, until the last few tasks where it becomes impossible to obtain a solution with the initially learned strategy. This way, during the experiment the participants had to switch their initial mental representation and encounter the associated negative transfer. Lastly, we conclude that a particular mental representation can enhance positive transfer in subsequent tasks that support this mental representation and create negative transfer in similar tasks that do not. Additionally, we observe that some mental representations are more available than others and facilitate stronger positive and negative transfer than others.

In our study, we provide existing information on mental representations and their decomposition, and connect it with positive and negative transfer, according to different views in psychology. In the chapter Introduction (1), we give a broad overview of the scientific field of problem-solving and its associations with mental representations. In the chapter Motivation (2), we explore the motivation behind our study. Later, in the chapter Theoretical background (3), we dive into each of the associated psychological concepts, examine their properties, and connect them in causal relations. The core of our study is the matchstick experiment, whose purpose is mentioned in the chapter Our study - aims and hypotheses (4) and whose design is described in detail in the chapter Methods (5). Afterwards, its results are analysed and discussed in the chapters Results (6) and Discussion (7), respectively. We wrap up this thesis in the chapter Conclusion (8) with some closing remarks on its future applications, possible improvements and interdisciplinarity.

Chapter 2

Motivation

The original idea behind this project was a simple "switch" of perspective that could lead to a better performance in certain problem-solving tasks. It describes how a simple element (usually an image or a problem) can be viewed and understood in radically different ways. Each of these ways of viewing the element, i.e. each perspective, exists and is real in itself, however, a person seems to always be limited to perceive only one at a time. An example might be found in the two perspectives of Figure 2.1, where either (1) a tree or (2) two animals can be seen, but only one at a time - seeing one prevents us from seeing the other. A perspective switch describes a change between perspectives and is an often-used tool in psychological puzzles and tricks, and in various Gestalt studies. But we are interested in its more measurable effects in everyday life - in problems and puzzles, and their respective solving processes.

Sometimes people are stuck with a problem and struggle with it for a long time, but when they switch their perspective of the problem, they can suddenly see the solution easily, which is often accompanied by shouting “Aha!” or “Eureka!”.

Why do we fall into these perception traps? And how does a perception switch help us solve such problems? These are some of the motivational questions that gave rise to our research study. During the course of the study, the term “perspective” evolved into a more precise and better understood term “mental representation” and our research focus narrowed to only one of the effects on the problem-solving process: the effect of transfer.

FIGURE 2.1: An image with two perspectives; one consists of a tree and the other of two mammals looking at each other

(Author: Rocio Castellanos, Title: Logo Intelligent, Collection: FIGURA FONDO, Image source: https://www.pinterest.com/pin/571464640192455484/)

Chapter 3

Theoretical background

In this section of our thesis, we investigate various concepts and ideas used in previous studies. Each of them is well described and supported with the experiments made by their authors. On the knowledge of these researchers, we build our own study.

3.1 Problem-solving in general

In general, problem-solving research studies individuals' practices used in identifying solutions to specified problems. This is a broad research field, as it encompasses many different problem types – from simple everyday problems (such as deciding what to have for dinner) and unconsciously tackled ones (such as when to take the next breath), to cognitively-demanding, complex problems (such as analysing chemical molecules for a vaccine) and unsolvable ones (such as disproving/proving the existence of a god). This field is also widely researched by many disciplines – most notably psychology, medicine, engineering, computer science, artificial intelligence and mathematics. Scholars often study problems in terms of their design, complexity, transparency, dynamics, and multiplicity of goals. All in an attempt to understand as much as possible about problems, so we can improve our performance and the performance of our technologies in an effort to find better and quicker solutions. Researching and understanding problems is in itself a problem that needs to be tackled.

One of the English definitions of the term ***problem*** is “a question raised for inquiry, consideration, or solution” (Merriam-Webster, 2020b). In this general way, a problem can be regarded as a difference between the actual situation and the desired situation. Another similar definition is made by Holyoak (1985): “A problem becomes to be when we see the goal but we cannot see how we could attain it”. A similar definition by Dunker: “A problem arises when a living creature has a goal but does not know how this goal is to be reached” (Duncker and Lees, 1945, p.1). A problem should have an initial state, a final goal, and the steps to achieve and reach the goal state (Mayer and Greeno, 1972). The term ***problem-solving*** or ***problem-solving process*** refers to the process by which problem-solvers reach a solution.

Although problem-solving is a very actively researched topic, one of its most important aspects often remains considered intuitive and very neglected, i.e. the effects of internal problem representation on the problem-solving process. Newell, Simon, et al. (1972), in their book Human Problem Solving introduced a theoretical framework for describing problem-solving. In their theory, problem-solving takes place in an external task environment, where out of the set of all possible internal representations, the problem-solver generates one (or more) problem representations and its associated search space. In this search space, the problem-solver performs practically oriented operations (often called “moves”) and with each operation moves through knowledge states from start to goal (Kotovsky and Fallside, 1989). We differentiate between internal and external search spaces. The internal problem space or representation space relates to the current problem representation and the associated mental models, and is defined by them. The external problem search space or task environment is calculable and describable; often, all possible operations can be identified and a mathematical tree or diagram can be created, mapping (all) possible moves and determining outcomes. For example, it is quite easy to create a tree of all possible moves and outcomes for the classic game of tic-tac-toe. This mathematical approach is studied by the mathematical subfield of game theory and can be an extremely useful tool in problem-solving. However, humans rarely dedicate the energy to mapping all possible operations and outcomes of each problem, and instead approach problem-solving in a more heuristic way. In our study, we are interested in human problem-solving and transfer, and will therefore focus our research on the internal problem space. In the following text, “problem space” will refer to the internal problem space and problem representations, unless specified otherwise.

As human minds, in contrast to computers, work in an associative way, they are not optimized to examine and search the complete problem space. To curb this limitation, we often resort to using a strategy and/or heuristic that can crucially affect the process and outcome of the problem-solving. This process has been called by Holyoak (1990, p. 271) "the acquisition of knowledge that restricts the need for extensive search". By employing a strategy or heuristic, the problem space is refined into a more familiar one, where the problem-solver already knows how to find a solution.

3.2 Problems versus puzzles

The terms *puzzle* and *problem* are often considered synonyms but, when examined closely, they differ in methodological nuances. To illustrate the differences, we define our problems and puzzles as follows: puzzles are tasks that have been specially designed by puzzle designers to have a particular solution, to be solvable in a particular way. Puzzle solvers are unaware of this solution and are tasked with finding it. Problems, on the other hand, are more organic tasks that have an uncertain number of solutions (from zero to infinity) and problem designers often have as much knowledge about the possible solutions as do the problem-solvers. The focus is shifted from the process of finding solutions in puzzles to an organic understanding of the problem elements and their interdependent interactions in problems. In problem-research, the focus is on how solvers approach and understand the task, compared to puzzle research, where researchers seek to find whether solvers are able to figure out the puzzle-designers' thought process and distil their designated solution.

3.3 Two schools of problem-solving

In the 20th century, most psychological articles that tackle problem-solving fall under one of the two schools of thought. These differ mainly in the aspects of and approaches to problem-solving that they investigate. In later chapters, we will focus on the Gestalt school, however, it is wise to briefly discuss the alternative school too.

3.3.1 Associationist view on problem-solving

Associationists have tried to explain the thinking process as an act of trial and error (Alzayat, 2011). Humans and animals tend to use certain strategies in a hierarchy by their commonality. We always initially attempt to solve a problem with the most common strategy and if this fails, we select the next strategy down the hierarchy. The original problem is split into smaller problems, each of which is tackled in sequence. Tackling/Solving a problem in the associationists view refers to applying different strategies to it in order to solve it. The main elements of this theory are: stimulus (a particular problem-solving situation), responses (a particular problem-solution behaviour) and the link or association between them. In other words, a problem stimulus which will stimulate a particular response during problem-solving. With each response, a participant learns what works and what does not, and the strategy hierarchy is updated.

A famous association theorist, Thorndike (1898), created the Cat in a Puzzle Box experiment. In this experiment, the cat must solve a puzzle by performing certain steps to get out of the box. His observation concluded that cats solved the problem by trial and error response at random. This experiment was later extended by Guthrie and Horton (1946), who conducted a more thorough study. They observed cats' successful trials and determined that the moves they make are almost identical for each trial. Their conclusion was that the cats' general plan does not change, but the order in which the strategies are attempted/tried does change between subsequent problems. This points to an existence of a hierarchy of strategies.

These strategies and behaviours are common to each particular problem-solver and do not differ between problems. Thus, in the associationist' view, it is irrelevant to distribute problems into different types.

3.3.2 Gestalt view on problem-solving

Gestaltists, on the other hand, focused their research on the underlining structure of the problem-solving process, the steps taken by problem-solvers, and the structure and properties of the problem itself. Their view of problem-solving consists of rearranging and manipulating parts of the problem (Alzayat, 2011). For every problem exists some form of a mental model with identifiable parts that can be mentally manipulated or changed. The process of mentally changing them is referred to as "thinking" or "problem-solving". While we participate in the problem-solving process, we compartmentalize aspects of a problem and relate them to one another, which results in an increased structural understanding of the problem. We gain the ability to comprehend how all parts of the problem fit together, how it is possible to reorganize them in a new way in order to satisfy the requirement and achieve the goal – to solve the problem (Mayer and Greeno, 1972).

In the world of Gestalt psychology, a ***mental representation*** or ***mental model*** is called a "***Gestalt***". The term "Gestalt" refers to an organized whole of elements and their interrelationships in a structure. This structure exists in any situation for an individual as an organized whole. In this thesis we will use the terms "mental representation" or "mental model" in place of "Gestalt".

3.3.3 Mental representations

Mental representations are indescribable perceptional entities that are specific to each individual. They are non-transferable between subjects and thus it is impossible to fully describe them using words. However, that does not mean it is not possible to study their effects on observable psychological aspects, such as behaviour or in our case problem-solving. In this aspect, they are similar to colours, which are inherently indescribable (it is impossible to describe the colour blue to a person who has never seen the colour blue), but there have been plenty of studies studying the effects of colours on a subject's emotional state (e.g. Strapparava and Özbal, 2010).

A mental representation of a problem is an internal image or a mental model of how the problem is perceived by a problem-solver. Mental representation is an important aspect of decision-making and thought processes. Different mental representations can shape the thoughts and actions taken by problem-solvers in completely different ways, although the problem is completely the same in every way.

Upon receiving a problem description, the subject's prior knowledge fully shapes the mental representation of the problem and, consequently, influences the problem-solving process as well. The effectiveness of the problem-solving process lies in the individual's perception of the problem elements and their available manipulations of these elements, which are integrated into the individual's mental representation of the problem (Kotovsky and Fallside, 1989).

A big part of mental representations is the process of creating or changing them, as this process has a significant and measurable effect on performance in a problem-solving process (Kotovsky and Fallside, 1989). In Gestalt psychology, a switch/change of a mental representation is often also referred to as ***restructuring***. The process of restructuring constitutes a change from one whole-view structure of the situation, not suitable for the task, to a different one (Derbentseva, 2007). There are a few more definitions of this term; Duncker and Lees (1945, p. 29) referred to it as a process where "parts of the situation which were formerly separated as parts of different wholes, or had no specific relation although parts of the same whole, may be united in one new whole". On another note, Ohlsson (1984) defined restructuring as "a change which affects the structural relations in the situation" (p.68) and "a process which changes the problem-solver's mental representation of the problem" (p. 71).

In the following sections, we will discuss the possible effects of different mental representations, especially the effect of their restructuring, on the process of problem-solving.

3.4 Insight in problem-solving

Research into insight began in the early twentieth century, mostly advanced by Gestalt psychologists. They separated problems based on the type of moves they require to reach a solution. We distinguish ***insight*** problems and ***non-insight*** problems; sometimes even hybrid problems, which are partially insight and partially non-insight problems.

Some researchers mark this as a continuous distribution, where problems can be assigned anywhere on a continuous spectrum between insight and non-insight extremes. Others treat this as a binary distribution, where the problem is marked as an insight problem if it displays any insight requirements. The latter distinction is more common, especially when problems are relatively small, such as problems used in psychological experiments that need to be compact and controlled as much as possible.

The term insight, originally defined by dictionaries as "the power or act of seeing into a situation" (Merriam-Webster, 2020a), is conceptualized in Gestalt psychology as "an act of restructuring the problem-solving situation which happens suddenly" (Derbentseva, 2007, p. 4). This definition encapsulates the suddenness, swiftness (of restructuring) as one of the main features of insight problems. Insight can also be understood as changing, reforming or restructuring a mental representation, as the problem-solver always ends up with another mental representation after insight occurs. Another definition by Mayer (1995) defines insight as "the process by which a problem-solver suddenly moves from a state of not knowing how to solve a problem to a state of knowing how to solve it" (Mayer, 1995, p. 3). An insight problem is a problem that requires the problem-solver to shift their perspective and view the problem in a novel way in order to reach the solution.

Ohlsson (1992, p. 4) describes the phenomenon of insight as a critical part of the process of solving insight problems with its role in solving an impasse. When a person attempts to identify or switch mental representations, which they have never done before, they experience an impasse. ***Impasse*** is a mental state when problem-solvers are "stuck" and unable to make any further progress in their ongoing problem-solving process. It is often accompanied by frustration and can be resolved by "overcoming it" with an insight (stemming from Ohlsson's definition). Overcoming an impasse is a time-consuming endeavour and leads to a delayed response, thus making it measurable and possible to study scientifically. Once a person overcomes an impasse, the mental representation becomes available and the person can make the switch easier, not needing to overcome such an impasse again.

Like Ohlsson (1992), another definition by Weisberg (1995) also defines insight problems by the impasse they create: "A pure insight problem can only be solved via restructuring". Simply put, they argue that an insight problem can only be solved by insight and that it contains an impasse. In the scope of our study we do not agree with this definition. Not every problem-solver will experience an impasse while solving an impasse problem. A problem-solver will have a high probability of not encountering an impasse if they solved the problem or a similar problem before and retained the generated mental model by overcoming that first impasse. While unlikely, it is also possible to generate the required mental model in the beginning of the problem-solving process. While solving insight problems, the majority of problem-solvers encounter some form of an impasse, but this cannot be generalized to all of them.

From a different perspective, which does not rely on an impasse, insight problems can also be defined by the steps that must be taken in order to solve them. These steps are not "locally rational" and directly leading to the solution (MacGregor, Ormerod, and Chronicle, 2001). Problem-solvers need to abandon the direct approach and find a "detour" (Köhler, 1925). An example how such a detour requires taking actions that do not lead directly towards the goal is shown in Figure 3.1. Another definition comes from Metcalfe and Wiebe (1987) who defined insight problems as characterized by a sudden discovery of a solution. Yet another definition came from Schooler, Fallshore, and Fiore (1995), who suggested that the difference between insight and non-insight problems lies in their main approach: insight problems are characterized by approach-recognition and non-insight problems by approach-execution. Grant and Spivey (2003), on the other hand, suggested that insight problems are defined by their solutions, which cannot be logically induced, as opposed to non-insight problems (like algebraic equations). All of these different definitions, that focus on various specific features, in their core describe the same common concept of an insight problem.

There are several degrees of difficulty for insight problems. Some are notoriously difficult, such as the famous nine-dot problem (Scheerer, 1963; MacGregor, Ormerod, and Chronicle, 2001; Kershaw and Ohlsson, 2004), which is unsolvable for the majority of problem-solvers without some additional info. The Two-Cord Problem (Maier, 1931) is generally solved by about half of the problem-solvers, while the Matchstick Problem (Knoblich et al., 1999) and the Tower of Hanoi (Kotovsky and Fallside, 1989) are solved by most people within a couple of minutes. Although they differ in difficulty, all of these problems, in order to be solvable, require a specific mental representation, which is not inherently obtained by an average problem-solver.

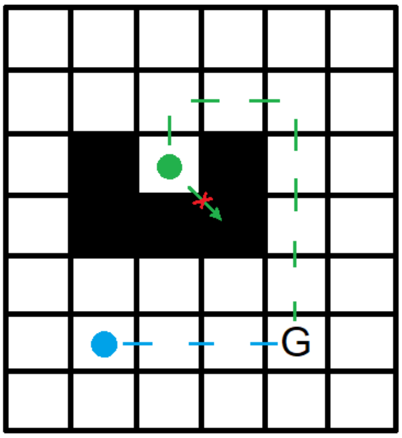


FIGURE 3.1: A detour requires taking moves that do not lead directly towards the goal.
The green path shows a detour, while the blue path takes solely direct moves towards the goal.

Insight is not limited to human problem-solvers. Köhler (1925) performed an experiment on chimpanzees, where they were given food that was out of reach. In one instance, the animals needed to connect together two sticks and, in another, they needed to stack some boxes in order to reach the food. Kohler also observed that the solutions to these problems were obtained in a quick, sudden manner.

We have observed multiple definitions of insight problems which define the same concept from different perspectives. As we are interested only in insight problems in this study, in the following sections of this thesis, the term ***problem*** will refer to insight problems, unless otherwise specified.

3.5 Productive and reproductive thinking

The process of thinking was, by Gestalt psychologists, split into two categories: "productive" and "reproductive" thinking (Wertheimer and Wertheimer, 1959; Mayer and Greeno, 1972). These terms are used to describe thought patterns used for creative and non-creative problem-solving. Wertheimer and Wertheimer (1959) in their book *Productive Thinking*, describe "productive" thinking as requiring an understanding of the intricacies of the situation, and “reproductive” thinking as the “blind” application of a previously learned approach.

Dominowski (1995), Wertheimer and Wertheimer (1959), and Luchins (1942) considered ***reproductive thinking*** a low-effort, habitual, mechanical thinking process, whose main problem-solving steps consist of reproducing steps that have already been successfully applied in previous problems. This type of thinking can only be achieved over a series of repeated similar problem-solving processes, where initially productive thinking had to be applied. With repetition, problem-solvers build their learned knowledge and habits, which then facilitate reproductive thinking. Compared to productive thinking, reproductive thinking requires lower cognitive effort and can sometimes be done completely subconsciously, requiring no cognitive effort at all. However, reproductive thinking is fixed to its existing routine and unable to deviate from it. Therefore, using reproductive thinking may prevent problem-solvers from discovering simpler, more effective solutions.

Productive thinking, on the other hand, is a demanding thinking process, focused on finding novel solutions. It is able to restructure the initial conceptualization of a problem, allowing the creation of new mental models where new strategies can be identified and applied. Productive thinking is often activated in an impasse; not only when reproductive thinking does not produce a result, but also when initially encountering a specific problem, which the problem-solver cannot solve by means of reproducing any of their known strategies. Changing an initially faulty mental model can break this impasse and give them more strategies to apply in their problem-solving process, and hopefully solve the problem.

Wertheimer and Wertheimer (1959) were researching how a general understanding of structural properties can improve and facilitate productive thinking. Students that were taught the process of calculating the area of a parallelogram were able to do so effectively, but were unable to calculate the area of a shape without straight edges. Another group of students, who were taught the structural properties of a parallelogram, was able to calculate its area and also apply this process to calculating the area of a different shape. Creating a more detailed mental model seems to facilitate strategy creation and improve productive thinking. Similar results have been found by Katona (1940) using card trick problems or matchstick problems.

3.6 Transfer

Mental models, in the Gestalt view, are hard to study scientifically. But this does not apply to studying their changes and the effects these changes have on the problem-solving process. A lack of these changes is associated with ***transfer*** (or learning). In essence, a problem-solver has identified a strategy and uses it as reproductive thinking in future problems. When this strategy is being applied to sufficiently similar problems (problem similarity is discussed later), problem-solvers do not need to generate the strategy anew (effectively skipping the incubation phase), and are able to solve these problems faster and easier. This process of successfully applying a learned strategy is called "***positive transfer***". In other cases, when this strategy is being applied to inherently different problems, problem-solvers reach an impasse, where the learned strategy cannot be applied to solve the problem. This process is adversely called "***negative transfer***" and it makes problem-solving longer, as problem-solvers firstly need to realize their solution does not work, switch to productive thinking, restructure their mental representation, overcome the impasse, and only then can they solve the problem.

Positive transfer can be observed in the study by Saugstad and Raaheim (1960), where participants were given a problem to move balls from a distance without touching them. They had a newspaper, strings, pliers, rubber bands and a nail at their disposal. The viable strategy was to bend the nail and attach it to a string. This hook needed to be thrown and used to pull the balls closer. Then the balls had to be caught in a rolled-up newspaper in the shape of a funnel, held together by the rubber band. Participants were split into two groups, where one of them was trained in object manipulation beforehand, such as creating a hook and a funnel. The participants, who had been given this training, were able to solve the puzzle in 95% of cases, while only 22% of participants without said training were able to so. Authors referred to this training as "making the function of the object available". This is a clear example of a positive transfer, which occurred in the participants that had already been acquainted with the strategy before solving the problem.

A similar example comes from the world of apes (Birch, 1945). Food was placed outside of the ape enclosure and the subjects (apes) were given a hoe, which could be used to reach the food. Only half of the participants were able to use the hoe to reach the food. For a couple of days, the subjects were given sticks to play with. Afterwards, the experiment was performed again and all the subjects were able to use the sticks and reach the food. They concluded that learning and positive transfer are not limited to human participants.

Negative transfer often occurs in problems which are similar on the surface, but differ in their core requirements. The surface similarity provides a false idea that an existing solution to one problem can be applied to another. An example of negative transfer is a study by Luchins (1942), who created an experiment where participants needed to solve the problem of measuring a specified amount of water. They were presented with 3 jugs of different sizes, an unlimited supply of water, and 11 problems. The first was a practice problem, used to introduce them to this problem type. The next five problems were puzzles solvable in a specific way; their purpose lay in introducing participants to a specific strategy and solidifying this strategy through repetition. The last five problems were solvable by the learned method and in an easier way by a different method. The control group was given just the introductory problem and the last five problems. More than 900 participants from various backgrounds participated in this study. The study found that most participants continued to solve the last five problems with the same strategy they had learned in the first five problems, compared to the control group which mostly found the quicker solution. This is a clear example of a negative transfer between the first five and the last five problems.

In this section, we discussed the Gestaltists' view in general, with an additional focus on insight, productive and reproductive thinking that form the essence of learning and transfer. In the next few sections, we will continue to discuss transfer itself in even greater detail. Afterwards, we will revisit productive thinking and its effects on breaking the transfer.

3.7 Analogical transfer

Conveying information is difficult, especially when this information does not have a physical shape, visual or auditory properties. And even with the help of those features, sometimes the receiver simply does not understand the information they are being conveyed. A common tool to improve information transition is an analogy.

> *"One good analogy is worth three hours discussion"*
>
> *– Dudley Field*

A piece of information is clarified/easier received if it is related and linked to existing information. Through association we use an existing, established and known concept, and associate its properties (physical, visual, functional, etc.) with a new concept we are trying to convey. Any differences can be added later for clarification, to differentiate between these concepts. A common example: "The structure of an atom is like a solar system. The nucleus is the sun, and electrons are the planets revolving around their sun. The atom is just a lot smaller." This example compares the difficult to understand concept of "an atom" with an already understood concept of "a solar system" (normally children learn about a solar system much earlier than about an atom), making it easier for the receiver to understand the received information. They do not have to make a completely new mental model, but can simply build upon the existing one (the solar system mental model) and expand it to cover the new concept (the atom).

It is important to note, that our description mentions concepts and aspects as opposed to mere physical objects. Analogies can expand mere objects and can be applied to whole situations and other more abstract concepts.

Analogical transfer is a transfer that occurs between subsequent problems, based on their perceived similarities. Wertheimer and Wertheimer (1959) studied this effect in children who had problems with understanding mathematical concepts and thus performed badly in school. When they were provided differently designed problems, which had a more direct relation to practical and real-world objects, they were more engaged, better understood the problems, and performed better in problem-solving, compared to the more abstract mathematical problems they had been struggling with earlier. Wrapping an abstract problem into a story, to which the problem-solver can relate, can have profound and significant effects on problem-solving.

In the analogical thinking process, we try to find the mapping of a relationship between two situations. Gick and Holyoak (1983) defined the process of analogical thinking as "the transfer of knowledge from one situation to another by the process of mapping – finding a set of one-to-one correspondents (often incomplete) between aspects of one body of information and aspects of another". Similarly, Gentner and Toupin (1986) defined analogical thinking as "the mapping of knowledge from one domain to another, as mapping from the base to the target problems". On the other hand, Holland et al. (1989) described analogical thinking as a central form of induction, which is used to generate inferences in situations that are deemed pragmatically important.

3.8 Analogical similarities

Analogical thinking and transfer are based on similarities, which can be split into two groups according to Davidson, Sternberg, and Sternberg (2003). The first group, ***surface similarities*** or silent similarities (as described by Vosniadou, 1989) are similarities based on the accessible components of the concept. These are easily accessible attributes, available through direct observation. The other group, called ***deep similarities*** or structural similarities, describe the core properties of a concept, more focused on the relational structure. Surface similarities are easier to notice, while deep similarities require some expertise.

An example of a surface similarity would be the tall wooden sticks used in both fences and scaffolding; an example of a deep similarity would be the purpose of separation that is inherent in both a fence and a wall. This separation into two groups is useful, as studies have shown that increased surface similarities between the target and the source problem enhance analogical transfer (Gentner, Landers, et al., 1985; Holyoak and Koh, 1987; Ross, 1984). In these studies, both positive and negative transfers were observed. In cases of positive transfer, deep similarities were also observed, while in cases of negative transfer there were no (or very few) deep similarities.

3.8.1 Steps of analogical transfer

Holyoak (1985) argues that analogical problem solving can be decomposed into four basic steps:

1. Constructing a mental representation of the already known (base) and the target problem.
2. Selection of the relevant source component to the target problem.
3. Mapping the base component to the target problem.
4. Extending the mapping to generalize a solution to the target problem.

When presented with a problem, problem-solvers first need to analyse and understand the problem, parse the relevant information, and create a mental representation. Afterwards, they select an analogous problem from their experience, for which they have already identified a solution. The strategy of this solution is then mapped and applied to the new problem. If the strategy used is successful in solving the new problem, we talk about a positive transfer; if not we talk about a negative transfer. Since it involves transfer, this kind of problem-solving is, of course, the reproductive kind.

3.8.2 Studies that support the transfer theory

A good example of the analogical transfer are the experiments done by Reed, Ernst, and Banerji (1974) called Jealous Husbands and Missionary-Cannibal problems. Both problems are spin-offs of the classic river crossing puzzle. In the missionary-cannibal problem, there are three missionaries and three cannibals who need to be transported across the river by a boat, but at no point in time can there be more cannibals than missionaries on any side of the river. And in the jealous husbands problem, there are three pairs of husbands and their respective wives, all of whom likewise need to be transported across the river by a boat, but in this scenario, a wife can never be left with other men and without her husband on any side of the river. These problems have a great surface similarity, but very different rules at their cores (low deep similarities). The study observed a significant negative transfer; because of surface similarities, a significant transfer occurred, but because of their differences, the learned strategy could not be applied and therefore this transfer actually negatively affected the problem-solving process.

Another such experiment was made by Kotovsky, Hayes, and Simon (1985). They used two variants of the Tower of Hanoi: (1) the Monsters and Balls Problem and (2) the Acrobats Problem. Both problems have the same essence, possible moves and their distribution, and the amount of required moves as the Tower of Hanoi. From the perspective of the mathematical discipline of game theory, all these three problems are identical. Their difference lies in their presentation. These problems are deep-similar but not surface-similar. The Monsters and Balls Problem consists of three different-sized monsters passing around three different balls with the restriction that a bigger monster cannot pass its ball to a smaller monster if the latter already holds another ball. The Acrobats Problem consists of three differently sized acrobats hanging from trapezes or from one another, with the restriction that smaller acrobats cannot hold their bigger counterparts. In this study, authors asked the participants to solve the Tower of Hanoi Problem and then one of the other two problems. They discovered that the Acrobats Problem facilitates much higher transfer rates than the Monsters and Balls Problem. Their argumentation for the observed data suggests that features of physical nature in the Acrobats Problem solidify problem rules and make them more available, therefore making participants less likely to make a wrong move. For example, the rule that a smaller acrobat cannot support the weight of a bigger one is inherently more available and makes more sense from a physical perspective than the rule that one monster cannot pass its ball to their smaller friend. Additionally, acrobats hanging from one another inherently block the bigger acrobats from moving. These perceived differences tie in neatly with the work of Hesse (2000) and his theory of analogy relations.

Another experiment by Kotovsky, Hayes, and Simon (1985) tested the initial difficulty of these problems. Participants were given three isomorphic problems: (1) the original Tower of Hanoi Problem, (2) the Monster Change Problem and (3) the Monster Move Problem. The latter two are variations of the above-mentioned Monsters and Balls Problem. In the Monster Move Problem, the monsters are passing the balls they are holding to each other and in the Monster Change Problem, the monsters are increasing or decreasing the sizes of these balls. They have much more similar mental representations compared to the Tower of Hanoi Problem, yet the minor differences between them still lead to a significant difference in problem-solving. The authors found that the Monster Change Problem took the longest, on average twice (2x) as long as the Monster Move Problem. And the latter took, on average, eight times (8x) as long as the original Tower of Hanoi Problem.

They assume that the differences originate in the different cognitive loads required to perform a move and in the different amounts of entities required to be held simultaneously in one's mind, as reported by the participants. They reached the conclusion that even slight changes in the problem design can have a significant effect on the resulting problem-solving performance.

Experiments called Peg Move and Dish Move, variants of the Tower of Hanoi, were made by Kotovsky, Hayes, and Simon (1985). In the Peg Move Experiment, participants were given three differently sized balls on three pegs or sticks and were asked to move them around following the standard Tower of Hanoi rules. The Dish Move Experiment was similar to the Peg Move Experiment, except that the balls were placed on three dishes instead of pegs, while the rules remained the same. Participants on average solved the Peg Move Problem in 160 seconds, while the Dish Move Problem took them on average 241 seconds. The authors relate the lower problem-solving performance with the increased informational load imposed by each problem. Just as the original Tower of Hanoi and Acrobats problems, the Peg Move Problem has an implicit feature that the top balls on a peg inherently block the movement of the bottom balls. The participant is thus less likely to consider such false moves and waste time with them.

Kotovsky, Hayes, and Simon (1985) went further and identified two phases of problem-solving: the Exploratory Phase (where the participant becomes familiar with the problem and its rules; here, transfer has the most effect) and the Final Path Phase (where the participant actually makes the required moves to solve the problem). Moves can be made in the Exploratory Phase but they are more random than not, as the participant is using them to become more familiar with the problem. The moves in the Final Path Phase are usually made at a much faster pace with a clear goal in mind. The authors also refer to the Final Path Phase as "a mad dash to a solution", because the average time per move is almost twice as short during the Final Path Phase compared to the Exploratory phase. In this study, the authors also created an experiment with more or less similar problems, which were all isomorphs of the Tower of Hanoi Problem. They came to the following conclusions:

1. Transfer happens in the Exploratory (move learning) Phase of the problem-solving process, while leaving the Final Path Phase of the process essentially unchanged. The participant has already become familiar with the

problem structure in the preceding "source" problem and can transfer this knowledge to the target problem. This skill acquired from the source problem should reduce the Exploratory Phase of the target process.

2. The amount of transfer was observed to positively correlate with the increased overlap in the representations of two problems.
3. The target of transfer is learning how to make moves.
4. The amount of transfer was observed to negatively correlate with the increased problem difficulty (measured by solution-time ratios). More difficult problems are deemed more resource-demanding and thus less able to facilitate transfer.

In this section we discussed the analogical transfer in detail. We have observed the importance analogies have on the effect of learning and transfer. We have found which types of elements and their properties facilitate analogical transfer in an either positive or negative manner. Analogical transfer was mostly attributed to surface similarities, which are also the most available and recognizable attributes. A change of these surface similarities can therefore significantly affect problem-solving performance.

3.9 Fixation & functional fixedness

Fixation is another commonly observed feature of insight problems. Scheerer (1963, p. 29) defined fixation as a process of "clinging misguidedly to a false premise or assumption concerning the task". It describes problem-solvers inability to change the current mental model. In its essence, fixation refers to a cognitive bias, which limits problem-solvers to only use objects in their learned, standard ways. It is an impaired creative ability to discover new, non-standard ways of object usage.

Scheerer (1963) exposed this concept in an experiment of triangles. Participants needed to form four equilateral triangles with six matches lying on the table. They were fairly unsuccessful, because they tried to form the triangles on a two-dimensional plane. The solution, however, can only be achieved in three dimensions.

Fixation is also referred to as functional fixedness by Duncker and Lees (1945). They describe it as a property of an object having a strong common function and not easily seeing it serving a different function. Additionally, learned functions and prior experiences that emphasize the usual function tend to inhibit the use of the object in a different, novel way.

Duncker and Lees (1945) studied this phenomenon with a Candle Box Experiment. In this experiment, participants were given tools and instructions to mount a candle onto a wall. One group of participants was given all the tools in a box and the other group was given each tool separately, including the box. The second group perceived the box as one of the tools and performed better at solving the problem than the first group, which perceived the box only as an initial container for the provided tools. Subsequent experiments showed that this effect occurs not only on physical objects, but also on mental objects and representations.

3.10 Einstellung (mental set) effect

The term "Einstellung effect" (sometimes also referred to as "mental set effect") was introduced by Gestalt psychologists and it means "setting" or "installation effect" in German. The term refers to participants' predisposition to solve subsequent problems in the same manner they had solved the preceding ones, even though better solutions might exist for subsequent problems.

Like fixation and functional fixedness, this is yet another phenomenon that hinders restructuring. "Einstellung refers to applying a previously learned rule or procedure to a task when there exists a simpler way of solving the task" (Derbentseva, 2007, p. 7). Derbentseva (2007) goes on to describe the Einstellung effect as "a blinding effect of habit". In this description, the Einstellung effect is equated with a negative transfer. In other literature, some authors take a broader approach and equate the Einstellung effect simply with "transfer". This way, the Einstellung effect is a negative or positive effect of past experience. In cases where the learned method leads to a solution, the Einstellung effect positively impacts the problem-solving process as the expected impasse never occurs. In other cases, it can enhance the impasse, making it more difficult for the problem-solver to change their mental representation. Luchins and Luchins (1950, p. 279) poetically described it as "instead of the individual mastering the habit, the habit masters the individual".

Luchins (1942), and later Luchins and Luchins (1950), have demonstrated the Einstellung effect with the renowned Water Jug Experiment, in which the participants needed to solve problems of measuring a certain amount of water using three jugs of different specific volumes. The participants were divided into two groups, where the control group did not partake in practice problems. The experiment concluded that in the tested problems, the participants in the experimental group tended to use the lengthier solution, which they had learned in the practice problems, despite the existence of a simpler one. Furthermore, one of the test problems was unsolvable by the learned solution. The participants in the experimental group had a difficult time solving the problem directly and generating a novel solution, thus indicating the presence of the Einstellung effect.

In a report by Blessing and Dronek (2006), Luchins (1942) argued that if the underlying problem-solving processes are well practised, problem-solvers can compose these steps in a single "mental set" and do not have to recreate them for every process. Blessing and Anderson (1996) argued that the underlying reason for the Einstellung effect is that problem-solvers, who had learned and practised a specific type of problems, can skip producing problem-solving steps altogether.

In the previous sections, we discussed several phenomena: positive and negative transfer, analogical transfer, reproductive thinking, fixation, functional fixedness, and the Einstellung effect. We argue the inherent similarity between all these concepts in a way that they all describe the same underlying phenomenon. They all refer to a tendency that problem-solvers attempt to solve problems in the same way they have solved previous similar problems, which can be beneficial or disadvantageous to their performance. In the following sections, we will mostly discuss the implications of our study on the concept of positive and negative transfer, however, note that all of our discussions can be similarly applied to and argued for any other concept.

3.11 Controlling the mental representations

Newell, Simon, et al. (1972) differentiated the two versions of the problem space: (1) the internal representation of the problem and (2) the external task environment. In this section, we will focus on the first version and explore its relation to the second one.

3.11.1 Hints

Additional information can also have a significant effect on transfer. It usually comes in the form of priming and hints, depending on whether it is given before or during the problem-solving process, respectively. Hints are notorious for their role in helping other problem-solvers with overcoming impasses. They are nothing more than additional information that is given to participants in order to direct them to the usage of a specific strategy or mental model to solve the problem at hand. The transfer rate of hints is huge, as their sole and widely understood purpose is to be applied directly and unchanged to the problem at hand.

The effect of hints was also measured by Davidson, Sternberg, and Sternberg (2003), where they analysed the effects of spontaneous transfer and informed transfer (the latter being attributed to hints). The group assigned to spontaneous transfer was given no hint and any transfer occurred solely by mapping the source problem to the target problem. The other group received an additional hint that helped them to map the source problem to the target problem. Davidson, Sternberg, and Sternberg (2003) argued that the positive informed transfer happened for one of two reasons: (1) participants were unable to access the relative analogy information in their memory, (2) participants were unable to apply the appropriate strategy they should have learned in the source problem. The hint helped participants to distinguish the strategy they needed to apply. In their study, participants were given two problems with same underlying core issue, but very few surface similarities. Participants without the preceding source/analogue problem were successful in solving the target problem in only 10% of the cases. If they solved the analogue problem beforehand, they were successful in 30% of the cases. And if they were presented with a hint, they were successful in 75% of the cases. In their study, hint has the highest success ratio, which is not surprising due to the nature of this additional information being given directly at the moment of an ongoing impasse and targeted directly at overcoming it.

3.11.2 Priming

While they both contain information that is meant to lead participants to the usage of a specific strategy, priming is quite different from hints. Priming is done on a participant's subconscious level through repetition and exposure to certain stimuli, and is generally done before problem-solving.

Kaplan and Simon (1990) suggest that problem-solvers almost always adopt the initial mental representation, that is suggested or hinted at by the provided problem description. They also argue that an adopted problem representation is never an issue until an impasse is encountered (Derbentseva, 2007).

> Kaplan and Simon's ideas give rise to an open question: Is a problem categorized as an insight or non-insight problem solely because of its description, which shapes the initial mental representation and creates the possibility for a future impasse? Would the problem not contain an inherent impasse if its description were worded differently?
>
> This question requires additional research before we are able to suggest an answer for it, because we simply do not understand the extent of the relationship between a problem and its description. Mainly, how much can we change the problem description without changing the problem? Any change in the problem description can have a significant effect on its understanding, perceived difficulty and initial mental representations. As we will discuss in the Size/Depth Experiment below, even a small difference in description can significantly change the problem.

Kotovsky, Hayes, and Simon (1985) created an experiment to test the importance of internal representation on transfer and they influenced this representation with priming. In order to research internal representations, all problems have to be exactly the same regarding their external representation. They constructed another two isomorphs of the Tower of Hanoi Problem, called the Size Problem and the Depth Problem. The problems were represented on a computer screen with the same screen containing three circles, representing different balls and each of them is positioned inside its own square. For the Size Problem, these balls were given the labels “small”, “medium” and “large”, and for the Depth Problem, these balls were given the labels “far”, “middle” and “near”. This sole external difference between the problems was made on the instruction sheet given to the participants, which contained the rules written using one or the other set of labels and the corresponding term "change the size" or "change the position". The effective problem rules, and the start and goal positions remained the same for both experiments. The same also applies to the visual changes, as a change in size and a change in depth can be represented by the same change on the computer screen. Therefore, the only independent variable in this experiment was the way the participant referred to the ball change in the experiment. We could argue that the participants were primed this way. The authors reasoned that this variation of how the ball change is referred to is a variation solely of the internal mental representation of the problem. In one variation of the experiment, some “windows” were added to the square to facilitate the illusion of depth.

The authors discovered that the target problems which were preceded by a source problem with the same representation were solved quicker than the target problems which were preceded by a source problem with a different representation. Their final conclusion was that it is the internal representation that determines transfer and that this representation can exist and operate somewhat independently of other problem features. Nonetheless, problems with similar features facilitate a bigger transfer. This conclusion makes internal representations empirically measurable, computer modellable, and scientifically understandable.

3.11.3 Incentives

Gathering participants for an experiment is an incredibly difficult task, as anyone who has ever done a large-scale experiment can attest to. Incentives are often used to attract participants, most commonly with monetary value. However, if the incentive is dependent on the participant's performance, it can have a severe effect on the latter.

According to Derbentseva (2007), several studies were made in order to investigate this phenomenon. Glucksberg (1962) performed Dunker's Candle Problem and found, that the participants in the group that got paid took longer to solve the problem compared to the group which was not paid. The author argued that incentives increased the motivation to stick with the initial mental representation and/or prolonged the extinction of this initial representation.

Similar results were reached by McGraw and McCullers (1979), who performed the Water Jug Experiment and found that the group that did not get paid solved the problem quicker compared to the group that got paid. In an analysis of these experiments, Camerer and Hogarth (1999) suggested that incentives led participants to exert more effort while working on those problems. This effort made them focus on the usage of familiar, already learned strategies of reproductive thinking. It stifled their productive problem-solving and led them to persist with the initial approach, thus prolonging their search for a solution. Reproductive thinking is commonly faster in performance, although not always applicable.

However, according to Derbentseva (2007) and Camerer and Hogarth (1999), the above-mentioned authors were unable to provide any significant evidence that problem-solvers tend to stick with the initial approach and instead attempt various unsuccessful approaches. On the other hand, Wieth and Burns (2006) found that the participants in their study who were given an incentive performed better than the control group. The topic of the effect of incentives on problem-solving still needs a lot of research, before we can come to some impactful conclusions.

3.11.4 Availability

Stemming from Simon's research, Kotovsky and Fallside (1989) argue that there exist inherent differences in mental representations, which result in their relative availability. ***Availability*** is a comparative measurement and represents the base difference between mental representations, that cannot be explained by transfer. The authors argue that "the availability of alternative representations explains differences in transfer. Furthermore, the availability of a representation is a determinant of the likely success of transfer" (Kotovsky and Fallside, 1989, p. 30). Simply put, some mental representations are inherently more available than others; they are easier to create, keep in mind, and switch to, when an impasse occurs. They are deemed less cognitively demanding.

Kotovsky and Fallside (1989) did not fully define availability and investigate its features and effects, but they did provide a good experiment, which neatly showcases this phenomenon. This is an investigative experiment, following the Size-Depth Problem experiment. They asked subjects, who had not participated in any prior experiments, to describe the Size-Depth Problem display. The majority of participants described the ball as changing in size, opposed to changing in depth. Afterwards, they were prompted by a researcher "that some people are able to see it differently", followed by a description of the ball changing in the other feature, such as "some see it (the ball) as changing in depth, going towards and away from the screen". This way, they were prompted to change their mental representation. Both times, participants were asked to report the intensity of the representation. From their responses, the authors concluded that size representation is more available than depth representation, and had a greater effect on the observed transfer; participants experienced more transfer between two subsequent Size Problems than between two subsequent Depth Problems.

3.12 Defying the transfer

In previous sections, we discussed in great length the broad and detailed workings of transfer and its effects on problem-solving. And as we have observed, transfer is both an aid and a hindrance. The moment transfer changes from having a positive effect on solving problems (positive transfer) to a negative one (negative transfer), the problem-solvers' best course of actions is to break the transfer and switch their mental models with productive thinking. In this section, we will discuss the moment when this happens in greater detail.

Negative transfer denotes that the transferred knowledge does not optimally apply to the given problem. Whenever we talk about negative transfer, we must have identified a better way of solving the problem, which is not in line with the transferred knowledge. This means that negative transfer will not always completely disrupt the problem-solving; sometimes, the problem-solving will just be carried out in a sub-optimal way. But more often than not, negative transfer simply cannot be applied to the problem to produce any viable solution. In this case, we talk about an impasse – a state where the transferred mental model is no longer useful.

3.12.1 On the occurrence of impasse

Why does the impasse occur in the first place? If the problem-solver is eventually able to solve the problem without learning any new knowledge or tampering with the problem, does this not mean that they had the ability to do so and all the necessary information all along? Why does the problem-solving process go through all the trouble of creating an impasse at all? According to Ohlsson (1992), the initial representation plays a crucial role in answering these questions. The initial representation of a problem creates a mental model, which defines a problem space as well as a space of potential solutions. If the new problem space does not contain a workable solution, an impasse will occur.

The concept of *insight* was defined by Ohlsson (1992, p. 12) as "the act of breaking out of impasse". The insight occurs as a result of change in problem representation, which is a result of productive thinking. Once the problem-solver has re-perceived the problem in a novel way, the new mental problem representation creates a new problem and solution spaces; and the new solution space might contain a solution to the problem.

Once an impasse has been solved - a new representation has been created - the impasse does not need to be broken again. Both the old and the new mental representations are available to the problem-solver and they are able to (with more or less effort) switch between them, and apply them to problems at hand. There is no need to create newer mental representations until another impasse occurs, where none of the existing mental representations are able to overcome it.

According to Köhler (1925), the problem-solvers (in his study: chimpanzees) who solved the problem themselves were able to repeat their obtained solutions when they were given the same problem again a week later. On the other hand, the problem-solvers who were merely shown the solution did not perform as well.

3.13 Constraint relaxation

Knoblich et al. (1999) and Ohlsson (1992) argue that knowledge consists of inherent constraints, which we build during a learning process. If the problem-solving process is understood as searching for a solution across the problem space, the constraints serve as boundaries of this problem space. An unobstructed search space in a problem-solving process would be too huge, and impossible to search properly. Constraints are useful in this regard, as they limit the possibilities the problem-solver considers and make the search manageable.

When a problem-solver encounters a problem that reminds them of a familiar problem, they will likely apply the constraints of that familiar problem to the new problem's representation. In this model, constraints are an integral part of knowledge and thus also subject to transfer. Initial mental models for new problems contain the same constraints as the mental models of problems they were transferred from. Knoblich et al. (1999) further argue that in order to overcome an impasse, some of those constraints need to be lifted or relaxed. In their model, forming a new mental representation is essentially a switch from one set of constraints, which defines an initial representation, to a new set of constraints, defining a new representation.

Now consider an example of a common problem of entering a room. One such common constraint for this problem is that doors should not be damaged during the process of opening them. In order to solve the problem of opening doors, the solution space demands that the problem-solver uses a key. Now consider a similar problem of entering a room in an emergency. In this situation, there is no time to search for a key and therefore no solution exists in this new mental representation. The problem-solver encounters an impasse. This impasse can be broken by relaxing the constraint that doors must not be damaged during the process of opening them. Without this constraint, the solution space expands and a solution to break down the door with an axe becomes a valid one. Going a step further, we can identify another constraint, namely that rooms should only be entered through a door. Assuming the room has no windows and thin, non-supportive walls, it is still possible to gain access to the room by breaking through a wall instead. There is virtually no limit on how many constraints can be identified and relaxed, each expanding the solution space in the process.

Constraints are separated based on their scopes within a given problem. The scope of a constraint is determined by how much a problem representation is affected when that constraint is relaxed. Wide constraints have a bigger impact on the problem representation than narrower ones. And the probability of constraint relaxation is directly proportionate to its scope in the problem space; constraints with a narrower scope have a higher probability of being relaxed, compared to constraints of a wider scope. Returning to our example of entering a room, the two problem constraints (1) doors and walls should not be damaged and (2) entry into a room is possible only through doors, have very different scopes. Constraint scopes are difficult to compare and normally require extensive experiments, but in our case the second constraint encapsulates the first one – in order to relax the second constraint, the first one also has to be relaxed. Therefore, the first constraint has a narrower scope and a higher probability of being relaxed in such emergency problem-solving, compared to the second one. It is much more common to think of the idea to break down a door than the idea to break down a wall.

In their Matchstick Experiments, Knoblich et al. (1999) introduced the idea of a relationship between the amount of change in the problem representation, as defined by the relaxation of constraints and chunk decomposition, and the problem difficulty. The difficulty of changing the initial mental representation to the required problem representation could be due to inappropriate constraints, which are imposed by prior knowledge and experience. Defining these constraints into separate difficulty scopes also allows us to define the relative difficulties of problem solutions. Relative solution difficulty depends on (a) the constraints that apply to the problem, (b) the scope of each constraint in relation to the problem and (c) the subset of minimal constraints that need to be relaxed in order to solve the problem. Everything else being equal, a solution that violates narrower constraints is easier to achieve than a solution that violates wider constraints; and a solution that violates fewer constraints is easier to achieve than a solution that violates several constraints.

Additionally, once a constraint is relaxed in the process of overcoming an impasse, it stays relaxed for any future impasses. The newly adapted mental model with the relaxed constraint can be observed to transfer to future problems.

3.14 Chunk decomposition

“Element of a problem” is a vague term and can apply to any part, object or subject, that is considered essential to the problem. Although the definition is vague, it is possible to argue that certain elements consist of other (smaller) elements in the problem. These recognized complex elements are referred to as “chunks”; the process of combining elements into chunks is called “chunking” and the process of splitting chunks into their respective elements is called “chunk decomposition”. In their nature, chunks and elements have their sources in the real world, but their effect is only relevant in their representation in mental models. In a mental model, the chunk and its core elements by definition have separate mental representations and do not normally exist simultaneously. For example, the problem of building a table can make use of separate mental representations for each leg, tabletop and drawers in its problem-solving process, while the problem-solving process of the different problem of cooking a meal makes use of a table in its entirety as a surface for food preparation. In the problem of cooking a meal it is not prudent to decompose a table into its core elements. Similarly, water molecules are a good representational model for studying molecular chemistry, but it makes little sense to use them as a representational model for studying waves and ocean tides.

Knoblich et al. (1999) argue that chunk decomposition is one of the mind’s responses to persistent failure in an impasse. If the current mental model cannot produce a solution, some chunks might be decomposed in order to restructure the mental model and expand the possible solution space. They further argue that the probability a specific chunk is decomposed during an impasse is directly proportionate with its tightness. Chunks can be defined as tight or loose based on whether their decompositions are themselves chunks. A chunk, which decomposes into other chunks, is loose and a chunk, whose decompositions cannot be regarded as chunks, is a tight chunk. An example of a loose chunk is a single sentence in the English language; it can be decomposed into words, which are themselves chunks. While a word is a tight chunk in this example, as it can be decomposed into letters, which are themselves not chunks. A single letter can technically be decomposed into separate lines (the letter "B" can be decomposed into one straight and two curved lines), but these lines by themselves hold no meaning in the context of a text.

Additionally, once a chunk type is decomposed in the process of overcoming an impasse, it stays decomposed. In other words, all such chunks are made more available to the problem-solver in their composed and decomposed version. The newly adapted mental model with decomposed chunks can be observed to transfer to future problems.

3.15 Matchstick task by Knoblich et al.

In support of the theories of constraint relaxation and chunk decomposition, Knoblich et al. (1999) conducted multiple experiments in which participants needed to solve simple matchstick equation problems. Participants were provided with mathematically incorrect (but valid) matchstick equations with Roman numerals and three different signs ("+", "-", "=", each denoting a respective arithmetic operation, i.e. addition, subtraction and equality), where each equation consisted of two signs and three numerals surrounding them. The participants needed to solve / correct them by moving one matchstick. The problems used in the experiments were separated into categories according to the relative difficulties regarding chunk- and constrain-types (discussed in the previous two sections).

They have identified three constraints that could be or needed to be relaxed:

(1) *Value constraint* – a numeral can be changed by moving a matchstick;

(2) *Operator constraint* – a sign can be changed by adding or removing a matchstick;

(3) *Tautology constraint* – participants are actually allowed to make tautology equations, where both signs represent equality and all numerals are the same.

On a separate note, they identified three separate chunks:

(a) *Tight chunks* – numerals (I, V and X) and the minus sign (-), which can be split into separate matchsticks, but these matchsticks have no inherent meaning by themselves;

(b) *Loose chunks* – numerals (II, IV, VII, IX, etc.), which can be split into two or more numerals (VII → V, I and I);

(c) *Intermediate chunks* – the plus and equals signs (+ and =), which can be split into other meaningful chunks, but are not commonly done so.

Overall, for example, a solution to the problem "VI = VII + I" is "VII = VI + I"; and to achieve this solution the participants needed to relax the value constraint and decompose a loose chunk.

Additionally, their work is supported by Alzayat (2011), who conducted a similar experiment. Participants were presented with four matchstick problems (with Roman numerals and addition, subtraction and equivalence operations) of the same perceived difficulty, and with one different problem, thus deemed more difficult. In the first four problems, the authors observed positive transfer and in the last one, negative transfer. Participants were separated into four groups, i.e. A1, A2, B1 and B2. Participants in groups A1 and A2 were given matchstick equations, where the solution was to move a matchstick between one Roman numeral to another, and participants in groups B1 and B2 were given matchstick equations, where the solution was to move a matchstick between an operator and a number or comparator (the only operator was always affected by the matchstick move). These different types of moves are referred to as strategies. In the first part of the experiment, each group was observed to experience positive transfer, as they solved the second, third and fourth problem significantly faster and with fewer mistakes after solving first problem. As for the fifth problem: participants in groups A1 and B1 were given a problem easily solvable by the strategy of the opposite group. Here a significant negative transfer was observed, as the participants needed more time and made more failed moves to solve the problem, simply because they had solved the four previous problems, where they learned a different strategy. The two remaining groups A2 and B2 were given a different fifth problem, where they needed to move a matchstick between the numbers five ("V") and ten ("X"). This problem also produced significant negative transfer for both groups, however, group A2 performed better (faster and with less failed moves) compared to group B2, because the strategy required for this problem was more similar to the transferred strategy of groups A1 and A2 compared to the transferred strategy of groups B1 and B2.

Findings by Knoblich et al. (1999) provided the following conclusions: the probability of relaxing a constraint positively correlates with the narrowness of the scope of the constraint. Similarly, the probability of decomposing a chunk is lower if tight chunks have to be decomposed than if only loose chunks need to be decomposed. Lastly, a conclusion about transfer: once relaxed, constraints stay relaxed and once decomposed, chunks stay decomposed. They concluded that constraint relaxation and chunk decomposition are significant determinants in task difficulty and transfer in the domain of matchstick arithmetic problems.

Chapter 4

Our study - aims and hypotheses

Our original motivation was an unanswered question: How to avoid perception traps, where the problem-solver inexplicably gets stuck and cannot succeed in finding a solution unless they change their view of the problem? This question is very broad and has been tackled in many studies we have mentioned in the previous chapter "Theoretical background (3)". Since this question is very broad, we decided to focus only on a small portion of it. In our study, we investigated the details of such perception traps. We created situations where the participant was inducted into seeing a problem in a certain way, and afterwards we created an impasse to force them into changing their perception of the problem.

4.1 Overview

The general aim of our study was to examine the effect of (participant's) learned strategies on the speed and efficiency of the problem-solving process and to what extent individuals tend to transfer previously learned strategies to subsequent problem-solving tasks. This transfer is an observable element of the underlying mental representation.

Using matchstick problem-solving tasks, we have investigated transfer in the case of two strategies (described in the section "Strategies (4.3)") and measured its effects in the amount of time, the number of moves, and the types of moves the participants took in the course of solving the problem. After learning the assigned initial strategy, the participants were tested on a set of matchstick tasks, starting with tasks that were solvable with the initial strategy and thus facilitated positive transfer, continuing with tasks that were less optimal, and finally ending with tasks that were completely unsolvable with the initial strategy, thus creating an impasse, demanding the participants to change their mental representation and facilitating negative transfer. The difficulty of learning a problem-solving strategy, of restructuring an associated mental representation, and of breaking an impasse is negatively correlated with the availability of the strategy and positively correlated with negative transfer and the difficulty of chunk decomposition and chunk relaxation (Knoblich et al., 1999).

The main source for our study regarding mental representations was the article by Kotovsky and Fallside (1989), in which they described *Herbert A. Simon's* work on transfer in problem-solving. We have already explained their study, which included the Tower of Hanoi Problem and its isomorphic tasks. Additionally, we expanded the work of Alzayat (2011) and Derbentseva (2007), who in turn based their work on the study by Knoblich et al. (1999), where participants needed to move matchsticks in order to solve arithmetic problems (puzzles) represented by Roman numerals with addition, subtraction, and equivalence operations.

4.2 Problem selection

Understanding how participants recognize and use different perspectives was quite a specific and deep research problem. A great disadvantage and restriction was the requirement of a sufficiently complex, yet malleable and manipulable set of problem tasks, which would fit this methodology. These tasks also needed to be commonly recognizable and affordable, so that they required no previous knowledge and experience in a specific field, which could otherwise lead to some biases.

For our experiment, we decided to use ***matchstick-equation problems***, which fulfilled the requirements outlined in the previous paragraph and relied only on primary-school-level arithmetic knowledge and familiarity with common arithmetic symbols. They were considered sufficiently simple and affordable for widespread public use – every person that met the experiment's age requirements was considered to be familiar with basic arithmetic and the symbols used. Matchstick-equation problems also have a lot of available manipulation options – the problems we ended up using, after restrictions had been applied, had a potential pool of 160,000 valid equation combinations. Lastly, another crucial element to keep in mind is that the way they are used in our experiment, matchstick problems are categorized as *insight* problems.

Matchstick-equation problems were already well established and well researched in the problem-solving-research community. Many studies had been conducted regarding them, most notably: Knoblich et al. (1999), Alzayat (2011) and Derbentseva (2007). To avoid replicating their studies, we decided to go a slightly different way and model our matchstick problems with the commonly known digital clock symbols (seven-segment display). Unlike Knoblich et al. (1999), we decided to forego manipulation of the comparator (equality sign) element and instead include an additional operator, making the equation slightly longer. Similarly to Derbentseva (2007), we decided to include operations such as "×" and "/", in addition to "+" and "−".

4.3 Strategies

Strategies are an integral part of a problem-solving process and shape its course. They specify their own subspaces of the problem-solving space and define the available moves. In our definition, they are each equated with its own mental representation and are thus transferable across tasks.

In our experiment, we tested participants on the following strategies:

> ***Strategy 'a'*** - moving a matchstick from one numeral to another numeral, or within one numeral;
>
> ***Strategy 'b'*** - moving a matchstick from one operator to another operator, or within one operator.

For each participant, at any given moment during the experiment each of these strategies was either *learned* or *non-learned*. The strategy was learned in one of two ways: (1) guided learning through the learning phase or (2) self-learning through insight.

(1) Learning a strategy through the learning phase is considered unavoidable – anyone participating in the learning phase learned the provided strategies which varied between groups. Strategies learned this way are referred to as *initial* strategies.

(2) Self-learning a strategy is done through insight – during the problem-solving process, the participant encountered an impasse and overcame it through insight (Ohlsson, 1992), with constraint relaxation, chunk decomposition (Knoblich et al., 1999), and restructuring of the mental representation. Strategies learned this way are referred to as *alternative* strategies.

Once a strategy was learned, the participant could always apply it in the scope of this experiment; however, it was not possible to do so before it had been learned. The term "to learn a strategy" is used to denote the restructuring of an associated mental representation and overcoming an impasse in the scope of this thesis.

Each of these strategies represents a particular mental representation. We were not interested in the actual mental representations the participants created in their minds, or in their phenomenological descriptions, but solely in the effect that the restructuring had on the problem-solving process. Actual mental representations could differ wildly among problem-solvers, but previous studies (Kotovsky and Fallside, 1989) strongly suggest that the effects of their restructuring on performance are significant, regardless.

With that in mind, we present here one of possible representations of the two strategies, just to provide an example for better clarity. In this example, each mental representation consists of two dimensions: an arithmetic one and a visual one. Each element used in the matchstick experiment (numeral, operator, comparator, and the equation itself) has its own visual dimension, which consists of one or several matchsticks distributed in some pattern on a surface. Additionally, each of these elements has an arithmetic dimension that gives the element an arithmetic value or function. For example, in the visual dimension the element "3" is a set of five carefully distributed matchsticks, while in the arithmetic dimension it is a numeral with the value of three. The restructuring of a mental representation and overcoming an impasse can be understood as switching from an arithmetic to a visual dimension – equivalent to switching from thinking about numbers and operators in an equation to the realization that those elements are made of matchsticks which can be moved around. Each of our two identified strategies corresponds to this impasse in either numeral or operator elements. Learning a strategy is overcoming the impasse for the corresponding element.

4.3.1 Constraints and chunks

Similarly to Knoblich et al. (1999), we have also identified constraints and chunks in our study. And since our study was similar to theirs, so were the constraints and chunks identified.
The following two constraints were identified:

(1) Numeral constraint

A numeral could be changed by moving a matchstick within a numeral or to another numeral. In our study, this constraint was related to strategy ‘a’. When strategy ‘a’ was learned, the value constraint was considered relaxed.

(2) Operator constraint

A sign could be changed by moving a matchstick within an operator or to another operator. In our study, this constraint was related to strategy ‘b’. When strategy ‘b’ was learned, the operator constraint was considered relaxed.

These constraints are similar to the first two constraints in the studies by Knoblich et al. (1999). But since we do not manipulate comparators, we cannot make one similar to the third constraint.

In this thesis we do not refer to these constraints specifically in favour of referring to strategies, as strategies and constraints are related: *Relaxing a constraint is related to learning a strategy*. Both of these strategies and constraints each have an effect on their own element type and since one cannot be said to have a greater effect on the problem representation, we argue that these strategies are of similar difficulty. Strategies 'a' and 'b' are orthogonal and learning one of them is not more or less difficult than learning the other.

Our chunks, on the other hand, are slightly different. In the scope of our experiment, there existed no numeral or operator that could be simply decomposed into two other elements. One could argue that from a single plus, one could make two minuses, but that would require additional rotation of the vertical matchstick. One single vertical matchstick had no inherent arithmetical meaning, since numeral 1 was made out of two vertical matchsticks. Therefore, every single numeral and operator was considered a tight chunk. They could be decomposed into separate matchsticks, but these matchsticks had no inherent meaning when standing by themselves and, moreover, could not be further decomposed. This way, we made no distinction between numeral and operator chunks. Similarly to constraints, once a participant learned to decompose a chunk of a specific type into its matchsticks, they were able to decompose any such chunk.

With these definitions, we have outlined one of the crucial design parts of our experiment: All tasks were of the same difficulty. And the only reason why one task was solved faster and/or with fewer moves could be assigned to positive or negative transfer.

4.4 Hypotheses

The findings of Kotovsky and Fallside (1989) show that the change of mental representation correlates with chunk decomposition, which can be effectively measured by a lowered performance in that particular problem-solving process. In line with their findings, we have outlined, in our own study design, that learning a strategy is necessitated by a change in mental representation. Learning a new strategy thus transitively leads to lowered performance in one of the tasks. Therefore, we expected the participants who had already learned both strategies and did not need to learn any other strategy would perform better than the other participants, who needed to learn at least one new strategy during the course of the experiment. Additionally, this expectation was supported by the authors' conclusion that once decomposed, chunks stay decomposed.

Therefore, we hypothesize:

> **Hypothesis 01 – Better with two strategies**
>
> Individuals that learn two strategies solve all tasks faster and more efficiently compared to participants that learn only one or no strategies.

We have learned from Kotovsky and Fallside (1989) and Davidson, Sternberg, and Sternberg (2003) that once problem-solvers have practised a problem and created a mental representation, they (subconsciously) attempted to apply this mental representation to future surface-similar problems. They were able to solve future deep-similar problems more efficiently, which resulted in positive transfer. On the other hand, the problems that were surface-similar but not deep-similar, resulted in negative transfer. These were the measurable effects of transfer, which could be measured alongside the general amount of transfer.

This effect was the clearest when only one mental representation was created and practised. When multiple different mental representations were created and available, problem-solvers were able to choose the appropriate one (switch between them) for the problem at hand much easier, and they did not have to generate a new one each time. This all leads to more flexible problem-solving, which was shown in a less strict transfer of a specific strategy. On the other hand, if a mental representation was not practised, it was less ingrained in the problem-solver's mind and the effect of transferring it was less apparent.

The effect of greater transfer on a single mental representation was additionally supported by Luchins (1942), who argued that if only a single mental representation was developed, problem-solvers tended to stick with it and use it in future surface-similar problems, regardless if using this mental representation actually hindered performance on these future problems or made them completely impossible to solve.

Therefore, we hypothesize:

> **Hypothesis 02 – Transfer**
>
> The transfer is greatest when only one particular problem-solving strategy is learned.

Stemming from the works of Kotovsky, Hayes, and Simon (1985), we expected the initial learning phase to have a positive effect on performance. In our experiment, the very first task the participants had to solve was solvable by the strategy they had just learned in the learning phase. That was true for all participants, except for the ones assigned to one of the control groups who had skipped the learning phase altogether. In order to solve this task, the participants, who had not yet learned any strategies, needed to create a viable mental representation during the problem-solving process of this task. Bearing in mind the cognitive difficulty in learning a new strategy and its associated decline in performance, we could reasonably expect these participants to perform worse than the other participants, who already knew the required strategy.

Therefore, we created an exploratory hypothesis:

Exploratory hypothesis 03 – Priming / Initial learning

Individuals that learn at least one strategy solve the first task faster than individuals that learn no strategies.

Additionally, we formulated a research question regarding the comparison of strategies used. We had conducted very little preliminary experimentation (with just a couple of testers) and could have significantly benefited from an additional general insight into the components of our experiment.

Therefore, we added a research question:

Research question 04 - Differences in strategies

Is one strategy more available (more commonly used) than the other? And does any strategy lead to a greater transfer than the other?

Chapter 5

Methods

In our experiment, the subjects were asked to solve simple matchstick tasks. In each task, they were presented with an invalid equation created by matchsticks, and were asked to move a as few matchsticks as possible to make the correct equation. They were also asked to do so as quickly as possible. We have identified two different mental representations in these tasks, each associated with one strategy of moving matchsticks. These strategies are defined in the section "Strategies (4.3)" and showcased in an example in Figure 5.1. By studying this experiment, we wanted to know how a participant, who has learned a strategy by solving a task, performs in a similar task that may or may not be optimally solvable by that strategy. In this thesis, the term ***task*** is used for a specific instance of a problem, that the participant has to solve.

Transfer in its essence was one of the primary focuses of our study. We investigated how priming, learning, and repetition solidified a specific strategy in the participant's mind and enabled positive transfer in the first few tasks given to them. Throughout the experiment, we gradually altered the tasks' optimal strategy and with it changed the transfer from positive to negative. The initial strategy became sub-optimal and eventually completely unusable (creating an impasse), making the participants experience (perhaps not consciously) a growing necessity for learning and switching to the alternative strategy. We observed at which point in the experiment the participants actually learned the alternative strategy and what effect this learning had on performance. The results were expected to show whether one of these strategies was more or less available/easier or harder to learn than the other, meaning the switch in mental representation (overcoming the impasse) was easier or harder to achieve, respectively.

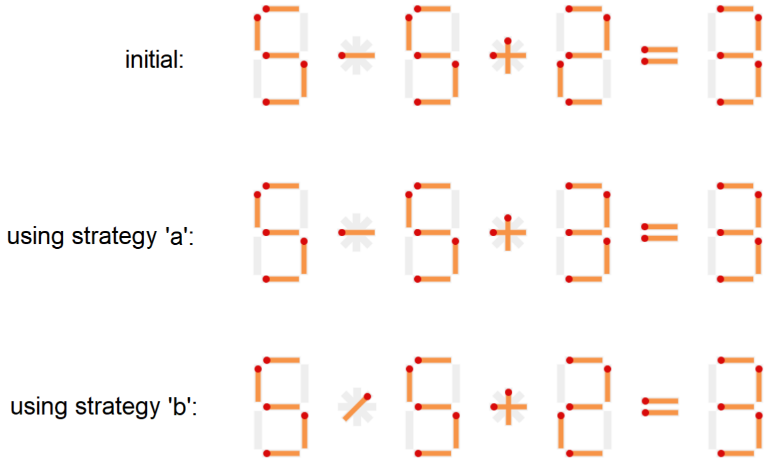


FIGURE 5.1: An example of an actually used initial equation (task iii1)
including one-move solutions with their respective strategies

5.1 Participants

Participants were adults between the ages of 18 and 68. Any person was actually allowed to partake in the experiment, but only participants who fit the age restriction were filtered and used for analysis. There was no perceived danger to participants' (mental or physical) health and participation was effectively anonymous. Participation was also completely voluntary with no reward or compensation. The participants were informed about the data being collected and only if they agreed, were they allowed to participate in the experiment. For more details on the agreement and the personal data we collected from them, see Appendix B - Sensitive information.

For the purpose of our study we required 120 participants (30 for each experimental group and 15 for each of the control groups). We performed this experiment on 254 participants. The number of participants in each group differed from the original design, due to some participants who had started but not finished the experiment, and thus slightly corrupted the group-assignment mechanism. However, since we had acquired more than enough participants for each group, as is evident in Table 5.1, it did not matter that some groups got more participants than others.

Groups	Required number of participants	Actual number of participants
A	30	70
B	30	57
0_a	15	36
0_b	15	36
0	30	72
AB_a	15	26
AB_b	15	29
AB	30	55
Total	120	254

TABLE 5.1: Number of participants assigned to each group

The majority of our participants came from Slovakia (61%), followed by Slovenian participants (23%) and a few Austrians (8%). The rest were well distributed across countries all over Europe and America. The participants were predominantly male (63%), followed by females (35%) and only 5 participants who opted for “Other sex”. Of the participants, 3% had PhDs, 28% were master’s degree holders, 35% were bachelor’s degree holders (or had a similar education level), and 30% were secondary school graduates. The mean participant age was 26.9 with a standard deviation of 9.1.

5.2 Materials

5.2.1 Independent variables

Not all participants experienced the experiment in the same way. Upon accepting the participation agreement, they were distributed into separate groups and each was given a set of problems to solve. The group they were distributed into represents the first independent variable, while the problems themselves represent the second. This section outlines and describes each of the two variables that we manipulated for each participant.

Strategy-learning groups

In the scope of our experiment, we introduced six **strategy-learning groups** as the first independent variable. These strategy-learning groups differed in two aspects, which are showcased in Table 5.2 alongside their possible values. Each participant was assigned to exactly one strategy-learning group.

	Possible values
Learned-strategy	$\{\ \emptyset,\ 'a',\ 'b',\ 'a'\,\&\,'b'\ \}$
Task-ordering	$\{\ (i) \rightarrow (v),\ (v) \rightarrow (i)\ \}$

TABLE 5.2: Aspects that were selected for each participant

A learned-strategy denotes the designated strategy that the participant learned in the first phase of the experiment, called the Learning Phase (discussed in the section "Learning phases (5.3.2)"). In this phase, each participant learned between none and both strategies, outlined in the section "Strategies (4.3)". Task-ordering builds on the learned strategy and denotes the order in which the task groups were attempted; participants could start with tasks solvable with strategy 'a' and continue towards tasks solvable with strategy 'b', or solve these tasks in the complete opposite order. Each participant always solved each task once.

The learned strategy and the task-ordering aspects were grouped together in six strategy-learning groups *{ A, B, 0_a, 0_b, AB_a, AB_b }*. These strategy-learning groups did not cover the whole spectrum of possibilities, as it made no sense that group A would attempt tasks starting with their alternative strategy (similarly for group B). The strategy-learning groups therefore formed the first manipulated independent *between-subject* variable.

Two control groups were used in order to account for learning none and both strategies. We added them following a suggestion by Alzayat (2011, p. 54) as an improvement on the studies of Alzayat (2011), Derbentseva (2007) and Knoblich et al. (1999), which also employed the use of matchstick experiments.

All the strategy-learning groups in our experiment were the following:

Group A – experimental group; learned strategy 'a';
tasks attempted in order (i) → (v)
Group B – experimental group; learned strategy 'b';
tasks attempted in order (v) → (i)
Group 0_a – control group; no strategy learned;
tasks attempted in the same order as group A
Group 0_b – control group; no strategy learned;
tasks attempted in the same order as group B
Group AB_a – control group; both strategies learned;
tasks attempted in the same order as group A
Group AB_b – control group; both strategies learned;
tasks attempted in the same order as group B

Participants were pseudo-randomly split into one of these groups. Each participant was randomly assigned into one of the groups with the lowest cumulative participant count. We adapted this calculation in favour of experimental groups, so that twice as many participants were assigned to them than to each control group. Since there were four control and two experimental groups, we ended up with half of the participants in experimental groups and the other half in control groups.

Tasks

The matchstick task (or **task**) was a mathematically incorrect/unsolved equation that was presented to the participant in order for them to solve it. The tasks could be tackled with matchstick moves and could have only one or multiple possible solutions, depending on additional requirements.

We required each participant, regardless of the strategy-learning group they had been placed into, to solve each task (the order, of course, differed based on their strategy-learning group); the **tasks** variable itself was the other manipulated independent *within-subject* variable.

Task groups

The tasks themselves were grouped in a task group and were solvable in the same way (regarding the strategies required). The tasks in a task group were always attempted in the same order regardless of how the task groups themselves were attempted. Keep in mind that task groups are not the same as strategy-learning groups.

The tasks were carefully selected, based on their possible solutions, and separated into five task groups:

Task group (i) – tasks are solvable only by strategy 'a'.
Task group (ii) – tasks are solvable by both strategies 'a' and 'b', but strategy 'a' is more efficient (the equation is solvable in 1 move only by strategy 'a' and in 2 moves by either strategy 'a' or 'b' or a combination of both strategies).
Task group (iii) – tasks are solvable by both strategies 'a' and 'b' equally efficiently (solvable in 1 move).
Task group (iv) – tasks are solvable by both strategies 'a' and 'b', but strategy 'b' is more efficient (the equation is solvable in 1 move only by strategy 'b' and in 2 moves by either strategy 'a' or 'b' or a combination of both strategies).
Task group (v) – tasks are solvable only by strategy 'b'.

Chosen equations

The number of tasks was set at *10*. We felt that this amount gave us sufficient data for the experiment, while still being manageable for the participants by not being too cognitively difficult.

With the help of custom software, specially written for this experiment, we were able to create a pool of all 160,000 possible equations for the selected equation frame. This was our starting pool, to which we applied additional task rules, making the pool smaller with each step. These rules and the process of applying them are discussed in depth in Appendix C - Technical design in the section "Equation selection (C.7)".

The resulting equations for the tasks were manually chosen from the pool of all possible valid, but mathematically incorrect/unsolved equations, which adhered to the additional task rules and the restrictions of their corresponding task groups. The chosen tasks are displayed in Table 5.3. Each task has a unique identifier, consisting of the task group identifier and the order number (written after the equation). The ten values of the "tasks" independent variable are therefore: *{ i1, i2, ii1, ii2, iii1, iii2, iv1, iv2, v1, v2 }*.

Task group (i)	2x3-3=5	*(task i1)*	5/3+2=3	*(task i2)*
Task group (ii)	5+5-5=3	*(task ii1)*	2/2*3=2	*(task ii2)*
Task group (iii)	5-5+2=3	*(task iii1)*	3*2/3=3	*(task iii2)*
Task group (iv)	3+3-5=5	*(task iv1)*	3*3-2=3	*(task iv2)*
Task group (v)	5/5+5=5	*(task v1)*	2/5+5=2	*(task v2)*

TABLE 5.3: Mathematically incorrect / unsolved equations used as tasks for participants to solve

There was a lot of fine-tuning that went into the design and selection of these equations. For more detailed information, see the complete Appendix C - Technical design.

In conclusion, the two independent variables that we manipulated in the course of our experiment were:

- Strategy-learning groups *(between-subject) – referred to also as* Groups;
- Tasks *(within-subject)*.

5.2.2 Experiment features

A few puzzles

In our experiment, we were interested in the organic thought process and applications of certain strategies and, therefore, wanted to focus on problems rather than puzzles; the difference between them is discussed in the section "Problems versus puzzles (3.2)". However, that is not always possible, especially in an experiment with an extensive requirement to control participants' mental states. As the core of our experiment, we used problem-type tasks, where the participant was provided with tools and was free to explore and tackle them in any way they chose, producing any valid solution. These tasks had been heavily researched and the majority of their possible solutions calculated before-hand. On the other hand, in the edge cases at the beginning and end of the experiment, the tasks had been restricted so much that they effectively became puzzles. In our experiment, we were bordering between problems and puzzles, restricting the freedom of problem-solving to better suit our experiment.

Complete surface similarity

Bearing in mind the research on analogical transfer by Kotovsky, Hayes, and Simon (1985) (which we have discussed in the section "Analogical similarities (3.8)"), we designed our own experiment to have complete surface similarity between problems/tasks that need to be solved, in order to facilitate transfer as much as possible. No surface differences had been identified, which could lead to transfer inhibition and introduce uncertainty in the experiment. In other words, at a glance, all the tasks in our experiment looked alike; they were quite similar equations, made out of matchsticks which needed to be moved in order to solve the task. All tasks followed the same principle and visual design, making the surface differences so minimal that they can effectively be disregarded. The sole differences between the tasks lay only in the deep structural similarities and differences (how the tasks were solvable based on strategy 'a' or strategy 'b') and we are certain that these similarities and differences are the cause of the positive and negative transfer which we were researching.

In our study, we focused on a very slight difference in mental representations – the difference between (1) the participant's initial mental representation and (2) the participant's mental representation with a learned strategy, decomposed chunks, and relaxed constraints. This was a very subtle difference - the smallest one we were able to identify. If we could prove the effect of such small mental representation changes, then we could extrapolate bigger effects on bigger changes.

Incentive aversion

Incentives in studies can be problematic according to Derbentseva (2007), as they can influence participants' performance in favour of reproductive problem-solving. This was discussed in depth in the section "Hints (3.11.1)". In our study, we had no incentives to offer our participants, so we are confident that this incentive bias had no effect on our study.

Hints aversion

As researched by Davidson, Sternberg, and Sternberg (2003), hints have an important effect on the problem-solving process, as they almost completely overcome an associated impasse, thus eliminating negative transfer. This was discussed in depth in section "Incentives (3.11.3)". In our study, we deferred from using hints in order to completely avoid any occurrence of informed transfer.

Because of this, quite a few participants found themselves stuck in an impasse, unable to proceed and continue with the experiment. These participants then (often angrily) ended their participation in the experiment and their collected data was discarded. This was one of the flaws in our experiment design that is discussed further in the section "Improvements (7.6)".

5.2.3 Demographic variables

- Age
- Sex
- Country
- Education

These demographic values have already been presented in the section "Participants (5.1)".

5.2.4 Dependent variables

Collected dependent variables

We collected the following four dependent variables for each of the test tasks separately. Thus, at the end of an experiment we ended up with forty collected dependent values, which we used in the analysis.

- Time – total time required to solve the task
- Moves – total moves made per task
- Specific_moves_a – total moves made in a task
  that correspond to strategy 'a'
- Specific_moves_b – total moves made in a task
  that correspond to strategy 'b'

Calculated dependent variables

During the course of the analysis, we calculated the variables given below. These were not additional dependent variables, but completely derived from the existing measured dependent variables.

- Ratio_of_moves_a – the percentage of all moves that correspond
  to strategy 'a' (for each task separately);
  also referred to as: *ratio of strategy 'a' usage*.
- Ratio_of_moves_b – the percentage of all moves that correspond
  to strategy 'b' (for each task separately);
  also referred to as: *ratio of strategy 'b' usage*.
- Decomposition_point – the number of the task in which the decomposition associated with the unlearned strategy happened. This variable is further discussed in the section "Hypothesis 02.03 - Negative transfer (6.2.5)".

Ratio_of_moves_a was calculated by dividing the *specific_moves_a* value with the *moves* value for each task separately. The result of this calculation was always a value between 0 and 1, which corresponded to the percentage of all moves made while solving a particular task that were made using strategy 'a'. Similarly, we calculated the *ratio_of_moves_b* with the use of *specific_moves_b* value.

Other important definitions

Joined groups

In the analysis part of this thesis, we have joined some groups and refer to them as joined groups. When performing an analysis with these group variables, we treated their participants and their data as if they were assigned to the same group. These groups are always compared only based on the ratio_of_moves values.

- Group 0 – joined groups 0_a and 0_b.
- Group AB – joined groups AB_a and AB_b.
- Experimental group – joined groups A and B.

Different task comparisons

- Standard task comparison
- Ordered task comparison

We also distinguish between task comparison. In most cases, we compare tasks of the same task-type together (*standard task* comparison) - we compare tasks that had the same unsolved starting equation, but might have been attempted in a different order (for example, task i2 was the second task for groups A, 0_a and AB_a, but the very last task for groups B, 0_b and AB_b). Aside from that, in a few cases, we compare the tasks in the order they appeared (*ordered task* comparison): the first tasks of every group are compared together, second tasks together, etc. - this can mean comparing different tasks for different strategy-learning groups (for example, the second task for groups A, 0_a and AB_a was task i2 and for groups B, 0_b and AB_b task v2). When the latter ordering is used for analysis, it is labelled "tasks in order".

5.3 Experiment design and procedures

5.3.1 Initial steps

Upon entering our website, the participant was anonymously registered and greeted by a language selection screen.

Right after the language selection and before any data was collected, the participants were redirected to an information page in their selected language, where a broad overview of the experiment was described, alongside the collected data and participation agreement. By clicking on a clearly marked button at the bottom, the participant agreed to continue and participate in the experiment.

After consent was given, the participants were interviewed for personal data (sex, age, their current country and their level of education). After all the data was obtained, the participants were directed to the task intro page, where they had the option to change the language or check out some basic information about their experimenters, before eventually continuing to the actual experiment.

After the experiment was completed, the participants were thanked for their service and given the option to provide a comment of their experience. These comments helped us to identify potential flaws and shortcomings or simply provided another perspective on the experiment. These comments are discussed in Appendix E - Possible improvements.

In the end, they were guided to the Familiar Figures Experiment, which was a part of a separate study and had no influence over the Matchstick Experiment. In this experiment, the participants were provided an image on the left side and six very similar images on the right side, one of which was an exact copy of the image on the left. They were asked to identify the exact copy as quickly as possible.

5.3.2 Experiment phases

The Matchstick Experiment consisted of eight distinct phases, which were always executed in the same order (some strategy-learning groups skipped some of the learning phases). In order from first to last, the phases were:

Phase L1 – Instructional Learning Phase
Phase L2 – Observational Learning Phase
Phase L3 – Practical Learning Phase
Phase T4 – Initial Strategy Required Testing Phase
Phase T5 – Initial Strategy Optimal Testing Phase
Phase T6 – Both Strategies Optimal Testing Phase
Phase T7 – Alternative Strategy Optimal Testing Phase
Phase T8 – Alternative Strategy Required Testing Phase

During the first three phases, the participant learned about the requirements of matchstick equation tasks and was familiarized with the equation style, its elements, and mechanisms of matchstick manipulation. In the first phase L1, the participant was provided with written instructions in their selected language. These instructions included a description of the task goal, a description of all elements on the page, and input options. In the second phase L2, the participant watched two or four videos of matchstick equations being solved. These videos included a mouse pointer and a visual blip, denoting a mouse-button click. In the third phase L3, the participant was given two or four simple matchstick equation tasks, which had to be solved in a very specific way. The tasks included arrows pointing to the matchstick that needed to be moved and the location to which it needed to be moved; the participant was restricted to perform this move exclusively. The above-mentioned equations, used in L2 and L3 were not used in the following testing tasks. This concluded the learning phase; afterwards, the participant was expected to have a full grasp on how to solve the experiment tasks.

In each of the following five testing phases, theparticipant solved two matchstick equations. The only difference among these phases were the equations provided and the goal instructions and restrictions.
In the first testing phase T4, the participant solved equations that were solvable only by their initial strategy. They could use exactly one move (if they made a wrong move, the move was reverted to the initial equation state). In the second testing phase T5, the participant solved equations that were solvable optimally (in one move) by their initial strategy and sub-optimally (in minimally two moves) by their alternative strategy. There was no restriction on the number of moves they could use. In the third testing phase T6, the

participant solved equations that were solvable optimally (in one move) by both their initial strategy and their alternative strategy. There was no restriction on the number of moves they could use. In the fourth testing phase T7, the participant solved equations that were solvable optimally (in one move) by their alternative strategy and sub-optimally (in minimally two moves) by their initial strategy. There was no restriction on the number of moves they could use. And in the last, i.e. fifth testing phase T8, the participant solved equations that were solvable only by their alternative strategy. They could use exactly one move (if they made a wrong move, the move was reverted to the initial equation state).

For an easier understanding of the experiment design, refer to Figure 5.2, which describes all the phases in order (from top to bottom) for each strategy-learning group separately. This information can also be supplemented with visual and functional design descriptions in the appendix section “Task page design (D.6)”.

Learning phases

In the process of learning matchstick task requirements and interactions with matchsticks, the participants were carefully guided with subtle priming hints during strategy learning.

The participants assigned to group 0 skipped the learning phases L2 and L3 completely; in learning phase L1, they were provided only bare information, without any priming examples. The following priming hints were applied only to other groups.

In the learning phase L1, the participants were primed to use their assigned strategy with a written example of a possible move: *“For instance, the matchstick tasks can be solved by moving matches from one number to another number or within that number.”* for group A, *“For instance, the matchstick tasks can be solved by moving matches from one operator (sign symbol) to another operator (sign symbol) or within that operator (sign symbol).”* for group B, and *“For instance, the matchstick tasks can be solved by moving matches from one number to another number or within that number or by moving matches from one operator (sign symbol) to another operator (sign symbol) or within that operator (sign symbol).”* for group AB. These written examples were visually assisted with a GIF video image of a matchstick being moved within a corresponding element; for group AB, both GIFs were shown.

In the learning phase L2, the participants in group A and group B were shown two videos solving matchstick equations with their learned strategy. The first video showed a solution with a matchstick being moved between same-type elements or within a corresponding element in the second video. The participants in group AB were shown all four above-mentioned videos in mixed order, starting with the both *between same-type elements* tasks and continuing with the *within a corresponding element* tasks.

And lastly, the learning phase L3 was designed similarly to phase L2, but instead of observing, the participants needed to solve the specified tasks. The participants in group A and group B were guided to solve two matchstick equations with their learned strategy. The first task was solvable by moving a matchstick between same-type elements, and the second task by moving it within a corresponding element. The participants in group AB partook in solving all four above-mentioned tasks, mixed in the same way as the videos.

In this way, the participants practised and learned their initial strategy, making it more available in subsequent testing tasks. We had not instructed them to use this specific strategy in any way. The alternative strategy simply required additional cognitive work to break the impasse of mental representations and make it available for use.

Testing phases

In the first testing phase T4, the participants were provided with a clear restriction that only one move could be made to solve the equation. The instruction stated *"Correct the equation by USING ONLY ONE MOVE"* with bold letters at the top of the page. In case they made a wrong move, all of the matchsticks were locked and could no longer be moved; the participant was clearly directed to click a button to reset the matchstick task to its initial state, effectively reverting the move. This move and the time it took to make it was still counted in the statistics data. The participants in groups A, 0_a & AB_a were given tasks from task group (i) and the participants in groups B, 0_b & AB_b from task group (v).

The second testing phase T5 lifted the restriction of only one move available. The participants were informed that they were able to use more moves, yet directed to solve the task using as few moves as possible, with a clear instruction *"Correct the equation by USING THE LEAST NUMBER OF MOVES"* written in bold letters at the top of the page. The participants in groups A, 0_a & AB_a were given tasks from task group (ii) and participants in groups B, 0_b & AB_b from task group (iv).

The testing phase T6 had the same instructions as phase T5. All participants were given tasks from task group (iii).

The fourth testing phase T7 was similar to the testing phase T5, only with switched task groups. That time, the participants in groups A, 0_a & AB_a were given tasks from task group (iv) and the participants in groups B, 0_b & AB_b from task group (ii).

The last testing phase T8 was similar to the testing phase T4, only with switched task groups. The one move only restriction and directive were reestablished. The participants in groups A, 0_a & AB_a were given tasks from task group (v) and the participants in groups B, 0_b & AB_b from task group (i).

5.3.3 Online aspect

Our experiment was conducted online. Since this format did not allow the experimenter to be personally available for participants during the data collection, all possible corner cases had to be identified beforehand and systems had to be put in place to guide the participants and prevent possible malicious intent. On the positive side, the digital distribution allowed us to reach a wider audience and thus potentially increase the number of participants for our study. The participants were able to participate in the study from the comfort of their home and multiple participants were able to perform the experiment at the same time. Additionally, each participant session was performed exactly as defined, eliminating all possible experimenter influence, and providing the ability for future scientists and researchers to examine the experiment design down to the smallest execution details.

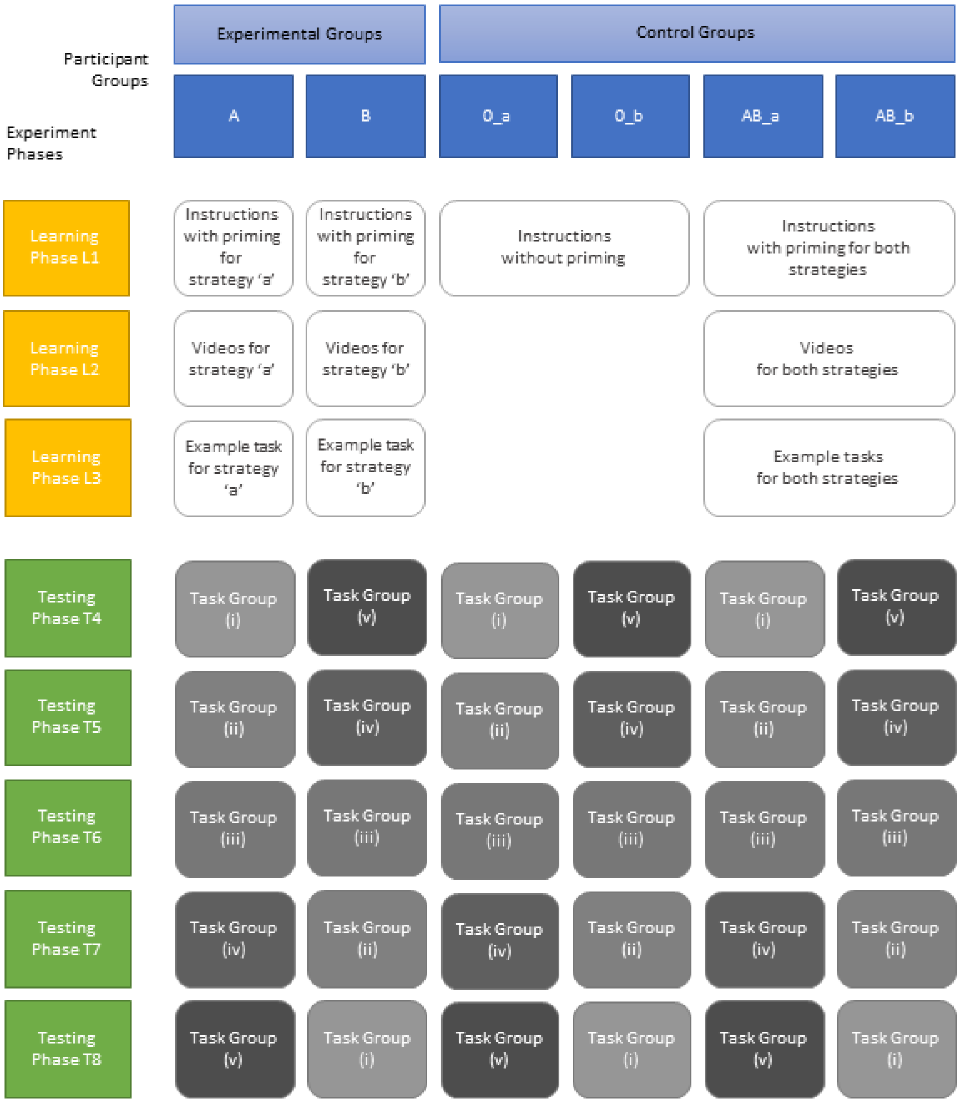


FIGURE 5.2: Visual representation of the experiment design

For more information on the design of the application, see Appendix D - Application design and for a more practical hands-on experience, participate in the experiment itself, available on the website `matchstick-task.eu`.

The experiment was conducted online between January and April of 2020. The experiment website was started and made available online, and potential participants were given invitations to participate via social media and other, more personal means of communication. The experiment website was left to run on its own and continue to collect data. When we determined that the amount of data was sufficient to perform the analysis, we extracted the data, but left the website to continue running for anyone who might be interested in participating in the experiment, without their data being used in the analysis.

5.3.4 Statistical analysis

In our experiment, participants were asked to perform tasks as quickly as possible and with as few moves as possible. This inherently meant that the dependent variables did not have a normal distribution, but instead a severely skewed right asymmetrical distribution. Figure 5.3 shows the distribution of one dependent variable.

This initial analysis showed us that the data was not normally distributed and we had to use non-parametric statistic tests, which were able to analyse non-normally distributed data. In the following analysis, we made use of the Mann-Whitney test, the Wilcoxon Signed-Rank test and the MANOVA.RM package (Friedrich, Konietschke, and Pauly, 2018) for programming language R version 3.6.3 (Team, 2017), which was able to perform ANOVA and MANOVA tests on non-normally distributed data with different sample sizes and heteroscedastic variances. We did not include Wald-type statistics, because they tend to be biased for small samples of non-normally distributed data (Randall, Arthur Woodward, and Bonett, 1997). To obtain significant ANOVA and MANOVA results, we also performed multivariate pairwise post-hoc tests (Tukey) between the groups with 95% confidence intervals. With these tests, we were able to fully and confidently analyse all the data required for the outlined hypotheses. In this thesis, the terms *ANOVA* and *MANOVA* refer to analytic tools from the MANOVA.RM package.

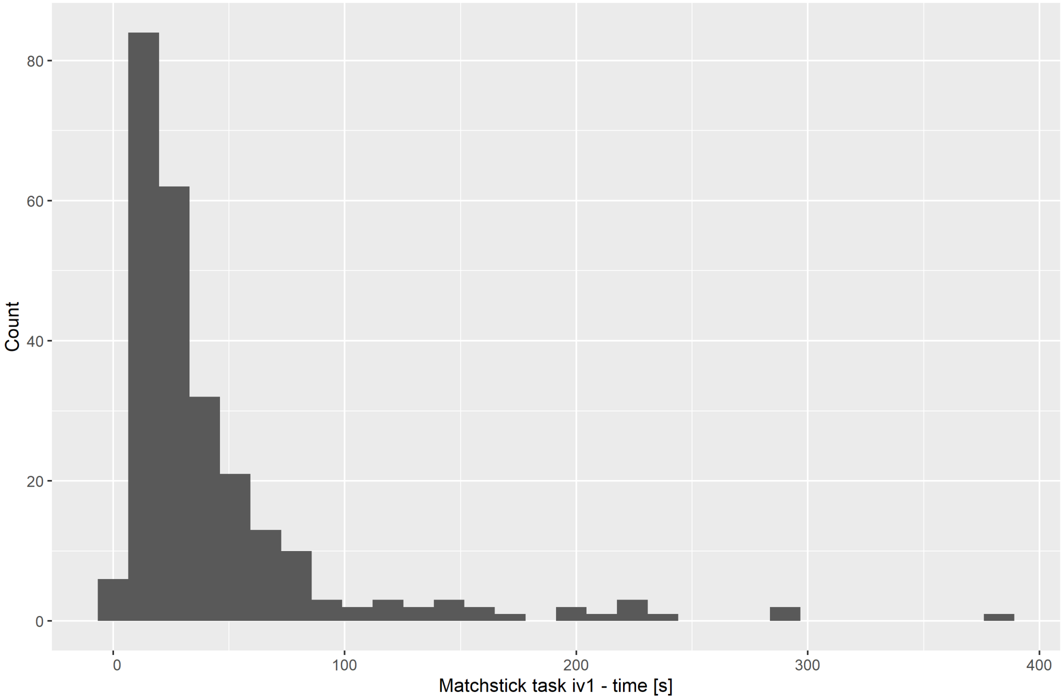


FIGURE 5.3: Raw time performance values for task iv

Initially we defined the significance with alpha; $\alpha = 0.05$. During the analyses, we conducted a lot of statistical tests, and although we do not mention it in the analyses every time, we adjusted the alpha value accordingly with Bonferroni correction for every analysis (this also applies to all marks of significance denoted by ‘ * ’ in tables and the analysis results).

Chapter 6

Results

6.1 Descriptive analysis

In our analysis, we could not compare tasks amongst themselves, since every task was unique and their difficulty might vary. These differences in difficulty could affect performance in terms of the time and moves required, but since we had no independent measurements of performance on each task, we were unable to determine these task difficulties to account for their effect. All of our measurements, except for the very first one, were affected by the transfer obtained from solving preceding tasks. Performance on each task, in terms of the time and moves required, could thus be fully attributed to *task difficulty* and *transfer*. Unlike task difficulty, it was possible to measure transfer. Task difficulty was deemed unique to every task, so by comparing different performances on a single particular task, we could eliminate the effect of difficulty and attribute those differences solely to transfer. In the section “Hypotheses testing (6.2)”, we always compare performance, in terms of the times and moves, only on the same tasks (the ratio of strategy usage can, however, still be compared among different tasks, since by design it accounts for performance differences).

In Figure 6.1 we can see differences in mean performances (times and moves separately) between tasks for all participants together, but we do not know yet how much of these differences were assignable to transfer and how much to inherent differences between tasks (task difficulty). An additional feature that we can observe in Figure 6.1 is that the times and moves are positively correlated for almost each task, confirming our assumptions that they serve as good measurements of performance.

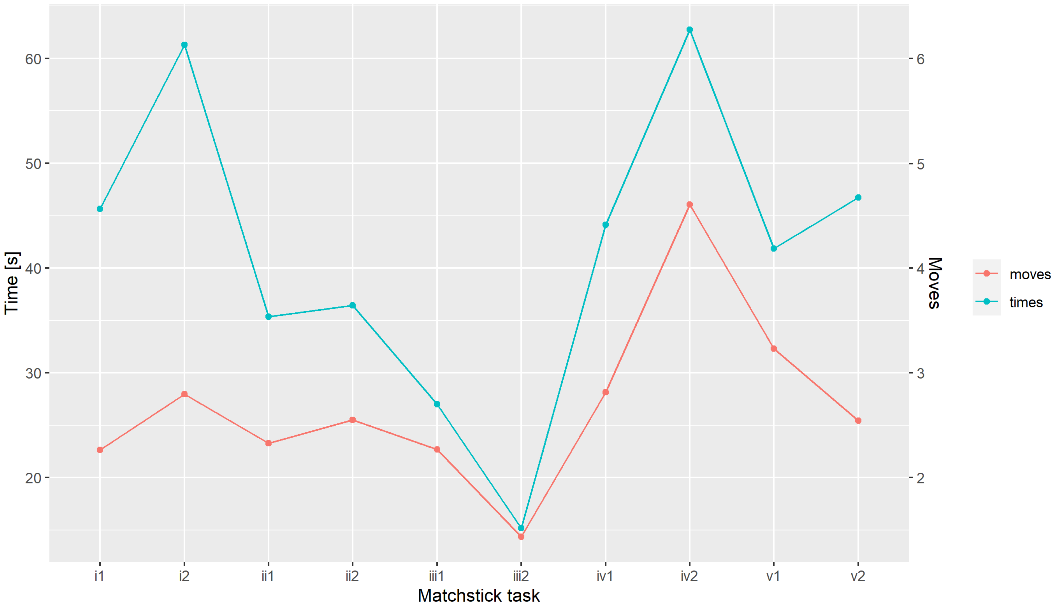


FIGURE 6.1: Mean performance (moves and times) in each task

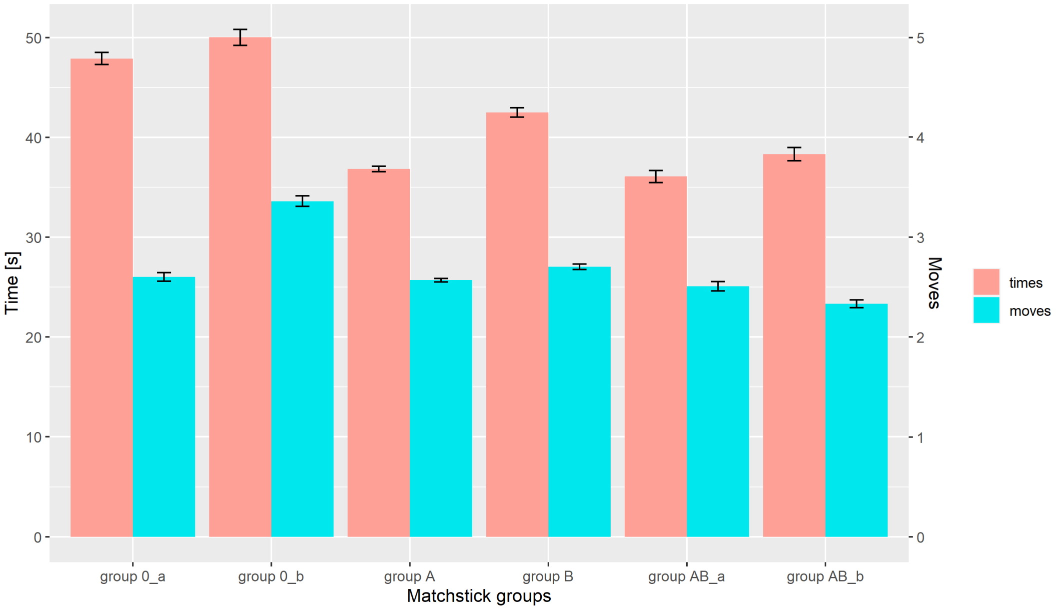


FIGURE 6.2: Combined performance (times and moves) on all tasks for each group separately

When comparing total performance in the experiment (the sum of all tasks), we found that groups indeed differed significantly amongst themselves, in regard of the time $[MATS\ \ Q_n = 558.811; p < .001]$ and moves $[MATS\ \ Q_n = 279.449; p < .001]$ values. This is also directly visible from Figure 6.2, where the combined/summed performance in times and moves is displayed for each group separately. In the subsequent hypotheses analysis, we analyse the groups and their performance in greater detail, in an effort to discover why these groups differ.

More detailed information about descriptive statistics can be found in Appendix A - Descriptive statistics, where the mean and standard deviation values are displayed for every group in every task and for every task in order. That data was used in most of the following hypotheses analyses and form the core of the graphs displayed in this chapter. Although we do not refer to it much in our discussion, it can provide a clearer and deeper understanding of our analyses for all who are interested.

6.2 Hypotheses testing

In this section, each hypothesis and research question are analysed individually and the results reported. Each hypothesis is split into several smaller working hypotheses, which outline the data used in the analysis and the requirements for the rejection of the null hypothesis (H_0). First, we attempt to reject each of the null hypotheses and afterwards we check whether the proposed hypothesis can be confirmed.

6.2.1 Hypothesis 01 – Better with two strategies

> Individuals that learn two strategies solve all tasks faster and more efficiently (with less time and fewer moves) compared to participants that learn only one or no strategies.

The following working hypotheses each compared one of the performance variables (the time variable or the moves variable) of the group AB (joined groups AB_a and AB_b), which had learned two strategies in the learning phase, against each of the groups 0, A and B, which had learned fewer than two strategies in the learning phase. The moves variables were compared only in tasks that did not restrict the number of moves a participant was allowed to make; therefore, we did not compare the moves in tasks belonging to task groups (i) and (v).

1. For every task, group AB spends less time compared to groups 0, A and B.
   (H_0: For every task, the group pairs AB - 0, AB - A and AB - B spend an equal amount of time.)
2. For every task of task groups (ii), (iii) and (iv), group AB makes fewer moves compared to groups 0, A and B.
   (H_0: For every task of task groups (ii), (iii) and (iv), group pairs AB - 0, AB - A and AB - B make an equal number of moves.)

Descriptive statistics are shown in Appendix A - Descriptive statistics in Tables A.1 and A.2 for the time and moves variables respectively. The analysis was conducted with ANOVA tests on each variable separately between groups 0, A, B, AB and the results are shown in Table 6.1. The only significant differences were observed in the time usage in tasks iii2 and v1. These significant differences were observed in only 2 out of 16 cases. We performed further analyses on these significant cases with pairwise comparisons; the results in the form of p-values are shown in Table 6.2.

The differences between groups in tasks iii2 and v1 could mainly be attributed to the differences between groups A and B. Even in the subsequent post-hoc analyses, we did not observe any significant difference between group AB and other groups. Thus, none of the null hypotheses could be rejected.

Even according to the graphs in Figure 6.3, it did not seem that group AB would have a significantly better performance (lower times or moves) in any experimental task, thus visually confirming our analysis.

Task	Time analysis	Moves analysis
i1	$MATS\ Q_n = 8.863; p = .040$	
i2	$MATS\ Q_n = 9.954; p = .025$	
ii1	$MATS\ Q_n = 12.198; p = .007$	$MATS\ Q_n = 10.483; p = .019$
ii2	$MATS\ Q_n = 12.098; p = .008$	$MATS\ Q_n = 7.777; p = .061$
iii1	$MATS\ Q_n = 5.552; p = .148$	$MATS\ Q_n = 3.488; p = .327$
iii2	$MATS\ Q_n = 16.712; p = .001$ *	$MATS\ Q_n = 5.910; p = .126$
iv1	$MATS\ Q_n = 1.785; p = .622$	$MATS\ Q_n = 3.361; p = .353$
iv2	$MATS\ Q_n = 12.364; p = .008$	$MATS\ Q_n = 4.906; p = .185$
v1	$MATS\ Q_n = 14.193; p = .005$ *	
v2	$MATS\ Q_n = 0.675; p = .880$	
α	0.005	0.0083

TABLE 6.1: Results of the ANOVA analysis of performance in terms of times and moves between groups 0, A, B, AB
('' - denotes significance)*

	task iii2	task v1
A - 0	0.4303	0.9997
AB - 0	0.9577	0.6983
B - 0	0.6206	0.4045
AB - A	0.3658	0.5084
B - A	0.3405	0.1619
B - AB	0.7875	0.7274

TABLE 6.2: p-values of the post-hoc pairwise comparison analysis of time usage in tasks iii2 and v1

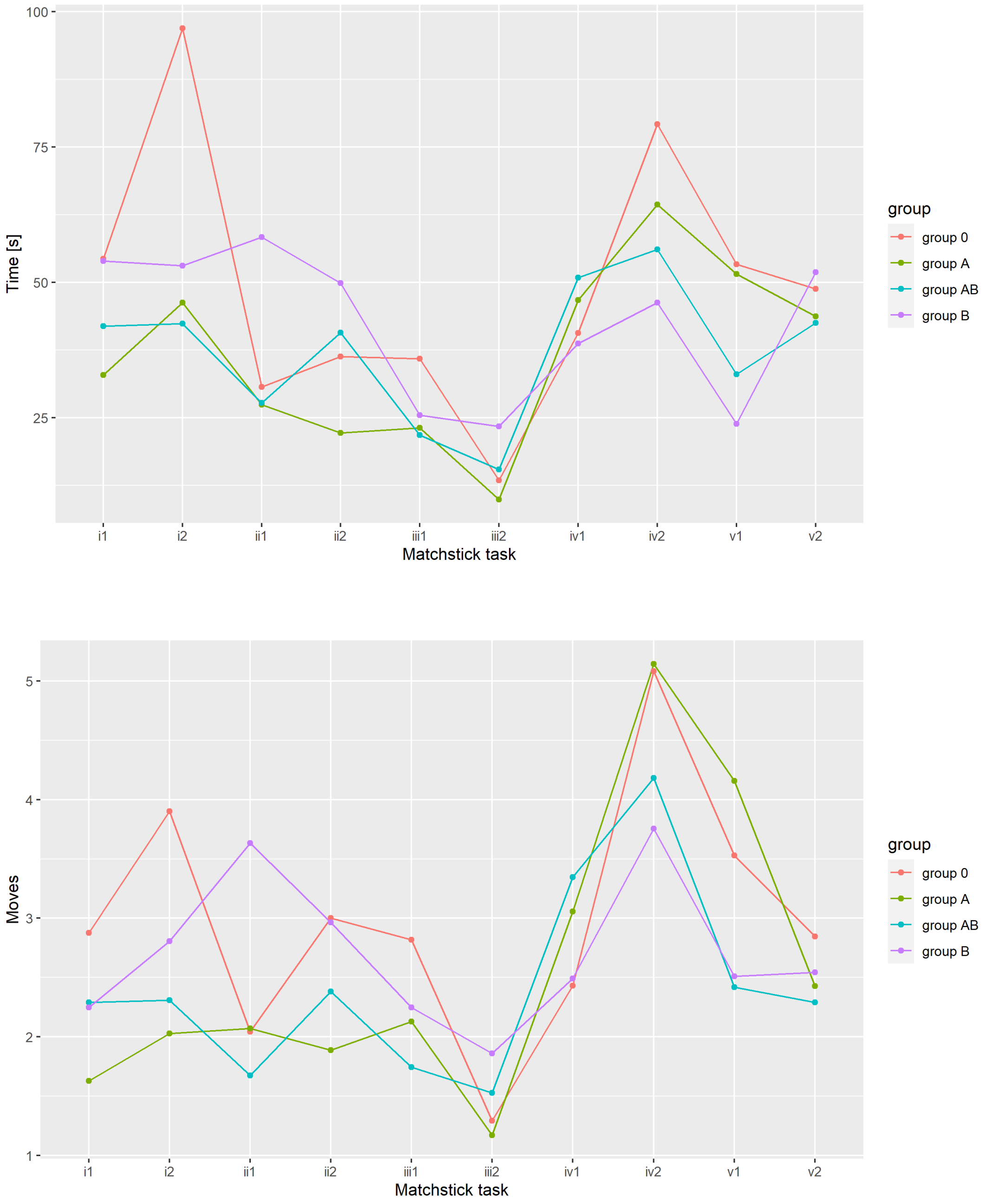


FIGURE 6.3: Performance in tasks between group AB and others (time - above, moves - below)

6.2.2 Hypothesis 02 – Transfer

> The transfer is greatest when only one particular problem-solving strategy is learned.

Groups A and B should have observed a higher transfer than other groups, which should have resulted from them using their initial strategy more often than the alternative strategy. This transfer should have resulted in a higher ratio of their initial strategy usage and subsequent performance impact.

For the purpose of analysis, this hypothesis was segmented into three separate sections, each of which focused on one aspect of transfer. In the first section, we focused on the actual measured amount of the transfer itself. In the second and third sections, we compared the performance impact based on the type of transfer – positive and negative, respectively.

6.2.3 Hypothesis 02.01 - Transfer ratio

The transfer ratio hypothesis expects the participants, who have learned and practised only one strategy, to use this particular strategy more often than other participants. Participants in group AB had learned both strategies and thus had the ability and incentive to use the most optimal strategy for each task. Participants in group 0 had not learned or practised any strategies, thus not having strengthened their mental representations in order to induce a strong transfer. Groups A and B were expected to have formed strong mental representations of their respective initial strategy and were expected to transfer this strategy in the testing tasks.

In the analysis of this hypothesis, we used move ratios which were calculated for each participant individually as a percentage of moves made corresponding to one of the strategies to all of the moves made in a particular task.

The following working hypotheses each compared the ratio values of one respective strategy usage in tasks that were solvable by both strategies. These values are calculated as described in the section "Calculated dependent variables (5.2.4)".

1. For every task of task group (ii), (iii) and (iv), the ratio of moves corresponding to strategy ‘a’ in group A is greater than the ratio of moves corresponding to strategy ‘a’ in groups 0_a and AB_a. *(H_0: For every task of task group (ii), (iii) and (iv), the ratio of moves corresponding to strategy ‘a’ is equal between group pairs A - 0_a and A - AB_a.)*
2. For every task of task group (ii), (iii) and (iv), the ratio of moves corresponding to strategy ‘b’ in group B is greater than the ratio of moves corresponding to strategy ‘b’ in groups 0_b and AB_b. *(H_0: For every task of task group (ii), (iii) and (iv), the ratio of moves corresponding to strategy ‘b’ is equal between group pairs B - 0_b and B - AB_b.)*

Descriptive statistics for ratio values are shown in Appendix A - Descriptive statistics in Table A.3. The analysis was conducted with ANOVA tests on each strategy separately. The results are shown in Table 6.3. We observed significant differences in tasks iii2 and iv1 regarding the ratio of the usage of strategy ‘b’. These significant differences were observed in only 2 out of 12 cases. We performed post-hoc analyses on these two significant tasks and displayed the results in Table 6.4. The differences observed were attributed mainly to differences between groups B and 0_b. While these differences were expected, they were only a fraction of the expected differences. Thus, we could not confidently reject the null hypotheses.

Task	**Strategy ‘a’**	**Strategy ‘b’**
ii1	$MATS\ Q_n = 0.885; p = .647$	$MATS\ Q_n = 4.994; p = .089$
ii2	$MATS\ Q_n = 0.765; p = .687$	$MATS\ Q_n = 5.738; p = .066$
iii1	$MATS\ Q_n = 0.184; p = .915$	$MATS\ Q_n = 36.172; p < .001$ *
iii2	$MATS\ Q_n = 1.640; p = .437$	$MATS\ Q_n = 2.323; p = .316$
iv1	$MATS\ Q_n = 3.913; p = .154$	$MATS\ Q_n = 8.254; p = .023$
iv2	$MATS\ Q_n = 4.236; p = .132$	$MATS\ Q_n = 12.874; p = .003$ *

TABLE 6.3: Results of the ANOVA analysis of the ratio of strategy ‘a’ between groups A, 0_a and AB_a and the ratio of strategy ‘b’ between groups B, 0_b and AB_b.
('' - denotes significance with α = 0.0083)*

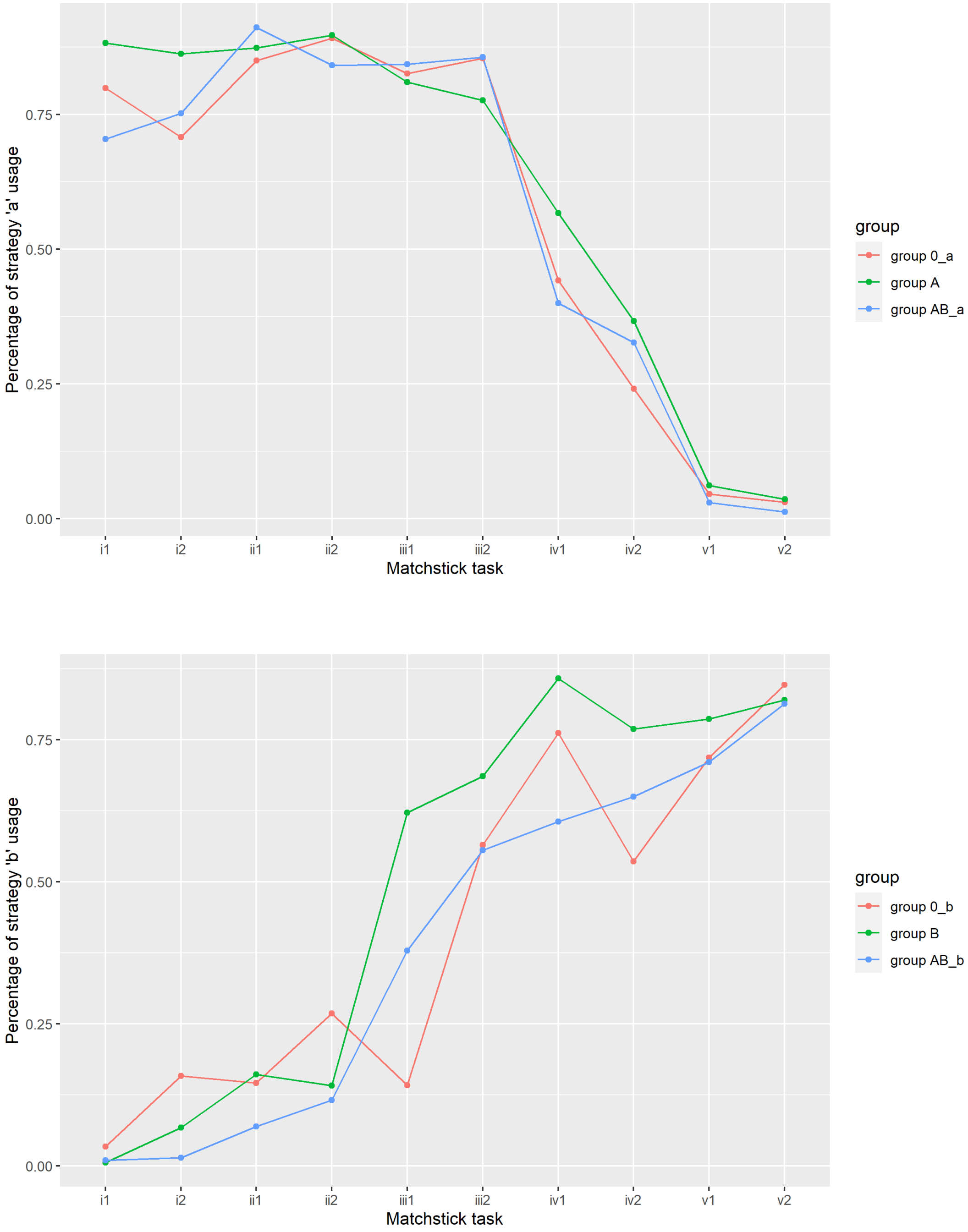


FIGURE 6.4: Percentage of strategy usage for every group with the same initial strategy
(strategy ‘a’ - above, strategy ‘b’ - below)

	task iii1	**task iv2**
AB_b - 0_b	0.1548	0.5081
B - 0_b	0.0001 *	0.0207
B - AB_b	0.1631	0.3517

TABLE 6.4: p-values of the post-hoc pairwise comparison analysis of time usage in tasks iii1 and iv2
('' - denotes significance with α = 0.0014)*

A visual inspection of Figure 6.4 provided a bit more insight into the results. For strategy 'b', the trend in the graph between groups B, 0_b and AB_b was pretty clear, as they follow a common slope. However, there was a slightly-significant difference in group B, which seemed to stick with their initial strategy more than the other groups. This was especially clear in comparison with group AB_b. But they were not so closely together as groups A, 0_a and AB_a. The graphs for these groups suggested far less difference between groups. These findings correspond with the analysis of Research question 04 - Differences in strategies (6.2.7), namely that strategy 'a' was more available and thus, the groups which knew strategy 'a' tended to prefer it to strategy 'b'.

And a final conclusion, contrary to the hypothesis: all groups with the same initial strategy exhibited a similar transfer.

For a better understanding, we expanded the analysis with an additional exploratory analysis, where we compared respective control and experimental groups to see if there was any difference in the strategy-usage ratio within them. We joined the data from groups 0_a and 0_b into control group 0, the data from groups AB_a and AB_b into control group AB, and from groups A and B into the experimental group. These groups were compared with regard to the ratio of each strategy usage. The results are shown in Table 6.5; for a better understanding, we added graphs in Figure 6.5.

According to the additional analysis on the ratio of strategy usage, there was a significant difference between strategy usage among experimental groups. More specifically, group A had a different ratio of strategy usage compared to group B. These differences were completely expected and attributed to a difference in transfer generated through the learning phase.

	Control group 0		Control group AB		Experimental group	
Comparison of strategy	a	b	a	b	a	b
Ratio in task ii1	0.001 *	0.000 *	0.259	0.034	0.062	0.000 *
Ratio in task ii2	0.044	0.073	0.710	0.763	0.000 *	0.000 *
Ratio in task iii1	0.000 *	0.000 *	0.000 *	0.000 *	0.000 *	0.000 *
Ratio in task iii2	0.074	0.967	0.003 *	0.001 *	0.000 *	0.000 *
Ratio in task iv1	0.101	0.356	0.015	0.010	0.000 *	0.000 *
Ratio in task iv2	0.267	0.182	0.408	0.497	0.000 *	0.000 *

TABLE 6.5: p-values of ANOVA analyses of the ratios of strategy usage within control group 0 (0_a and 0_b), control group AB (AB_a and AB_b) and experimental groups (A and B), respectively ('*' - marks significance with $\alpha = 0.0083$)

There were also significant differences in the strategy ratios in some of the tasks between the pairs of groups 0_a - 0_b and AB_a - AB_b. We expected some differences between them, but not significant ones. The results suggest that these control groups show higher amounts of transfer than we originally thought. They more closely resembled their corresponding experimental groups than their respective other-initial-strategy control groups.

In other words, group 0_a more closely resembled the ratio of strategy usages of groups A and AB_a than the ratio of strategy usage of group 0_b. Similarly, group AB_a more closely resembled the ratio of strategy usage of group 0_a than it did group AB_b. Respectively, groups with initial strategy 'b' more closely resembled the ratios of strategy usage of other groups with initial strategy 'b' than they did other respective control or experimental groups. The graphs in Figure 6.5 suggest greater discrepancies and a higher significance, so we expect that ANOVA would provide the same higher amount of measurements.

In conclusion, there seem to be a greater similarity in the ratios of strategy usage among the control groups than was initially expected. These similarities could easily be attributed to an unwanted transfer amongst control groups. In further analyses, we will determine how much of an effect this feature has on our experiment.

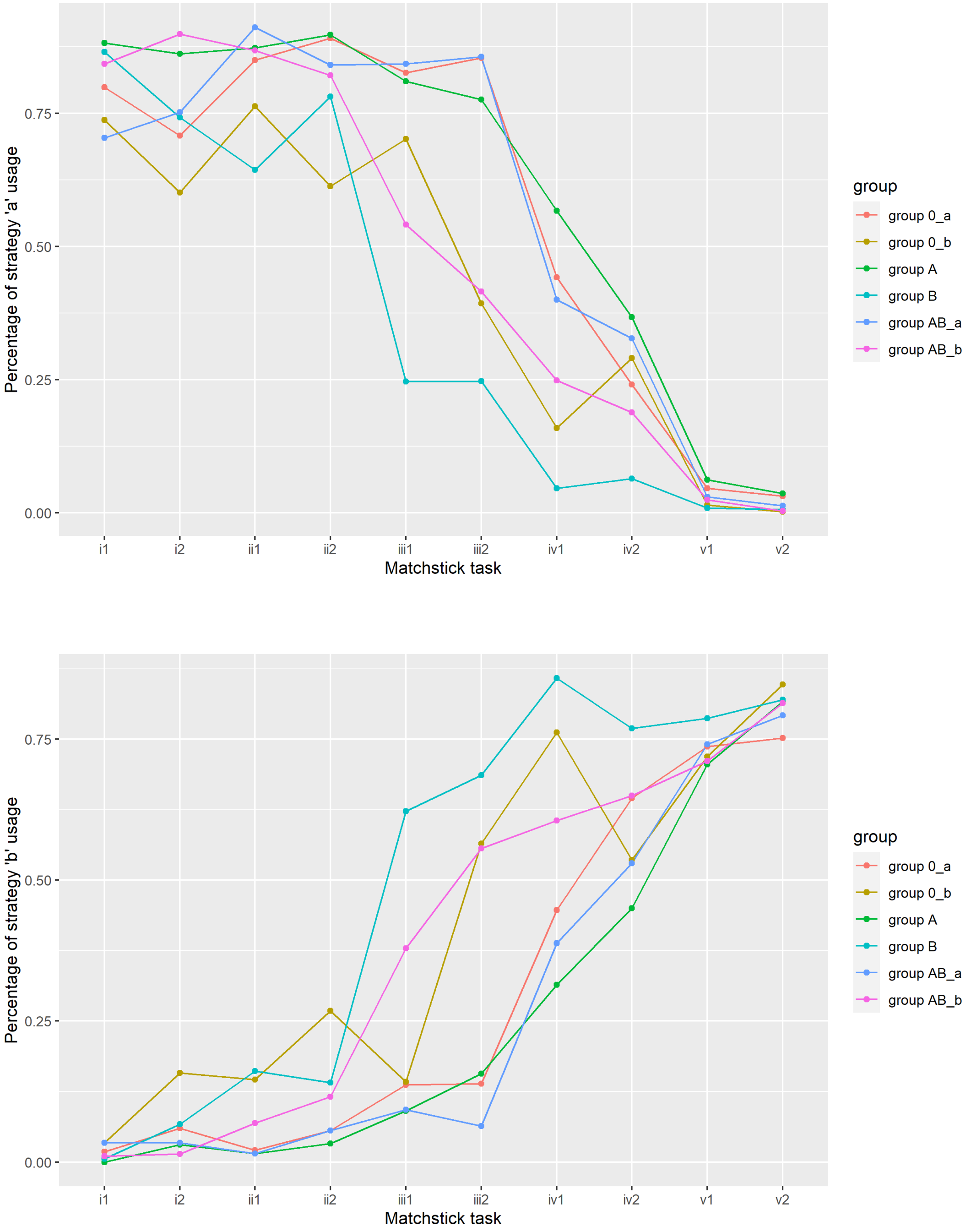


FIGURE 6.5: Percentage of strategy usage for every group (strategy 'a' - above, strategy 'b' - below)

6.2.4 Hypothesis 02.02 - Positive transfer

According to Davidson, Sternberg, and Sternberg (2003), transfer in deep-similar problems should lead to improved performance – positive transfer. The positive transfer hypothesis thus expected participants to perform better (with less time spent and fewer moves made) in problems that were optimally solvable with the initially learned strategy. This applied to the first six test tasks in order for each group.

We separately analysed the performance of groups related to each initial strategy, and separately for each of the performance variables. Notice that tasks 1 and 2 were restricted to one move only, thus we could compare performance only with regard to the time variable.

1. For every task 1 – 6 in order, group A spends less time compared to groups 0_a and AB_a.
   (H_0: For every task 1 – 6 in order, the group pairs A - 0_a and A - AB-a spend an equal amount of time.)
2. For every task 3 – 6 in order, group A makes fewer moves compared to groups 0_a and AB_a.
   (H_0: For every task 3 – 6 in order, the group pairs A - 0_a and A - AB_a make an equal number of moves.)
3. For every task 1 – 6 in order, group B spends less time compared to groups 0_b and AB_b.
   (H_0: For every task 1 – 6 in order, the group pairs B - 0_b and B - AB_b spend an equal amount of time.)
4. For every task 3 – 6 in order, group B makes fewer moves compared to groups 0_b and AB_b.
   (H_0: For every task 3 – 6 in order, the group pairs A - 0_b and A - AB_b make an equal number of moves.)

Descriptive statistics for ratio values are shown in Appendix A - Descriptive statistics in Table A.3. The analysis was conducted with several ANOVAs for each of the strategies and variable combinations separately. The results of performance between groups with different initial strategies are shown in Table 6.6 with the p-values separated according to the strategy and variable used in the comparison. The only significant differences were observed in the time variables in task i2 for initial strategy 'a' and task iv1 for initial strategy 'b'. These two were the only significant differences out of 20 tests, which could be attributed to within-subject variance. Thus, we were unable to reject any of the null hypotheses.

	initial strategy 'a'		initial strategy 'b'	
Task	time	moves	time	moves
i1	0.010			
i2	0.006 *			
ii1	0.455	0.729		
ii2	0.616	0.941		
iii1	0.312	0.299	0.354	0.680
iii2	0.169	0.363	0.029	0.133
iv1			0.002 *	0.059
iv2			0.564	0.572
v1			0.657	
v2			0.112	
α	0.008	0.012	0.008	0.012

TABLE 6.6: p-values of ANOVA analyses of performances amongst groups with initial strategy 'a' (A, 0_a and AB_a) and initial strategy 'b' (groups B, 0_b and AB_b) ('*' - marks significance)

No significant differences, which would be in favour of a positive transfer in the experimental groups only, could be discerned from the ANOVAs. The graphs in Figure 6.6 similarly suggest that group A did not perform significantly better (in terms of time and moves usage) than other groups with initial strategy 'a' in any tasks. A similar statement could be made for group B and other groups with initial strategy 'b', although we have omitted the graphs of groups with initial strategy 'b' for brevity.

However, according to the graphs, the improvement in performance in subsequent tasks, which were optimally solvable by their initial strategy, was clear for both strategies. This result was expected and could be solely attributed to the effect of positive transfer. This supported the findings of the analysis of hypothesis 02.01, which discovered some effects of transfer within control groups as well.

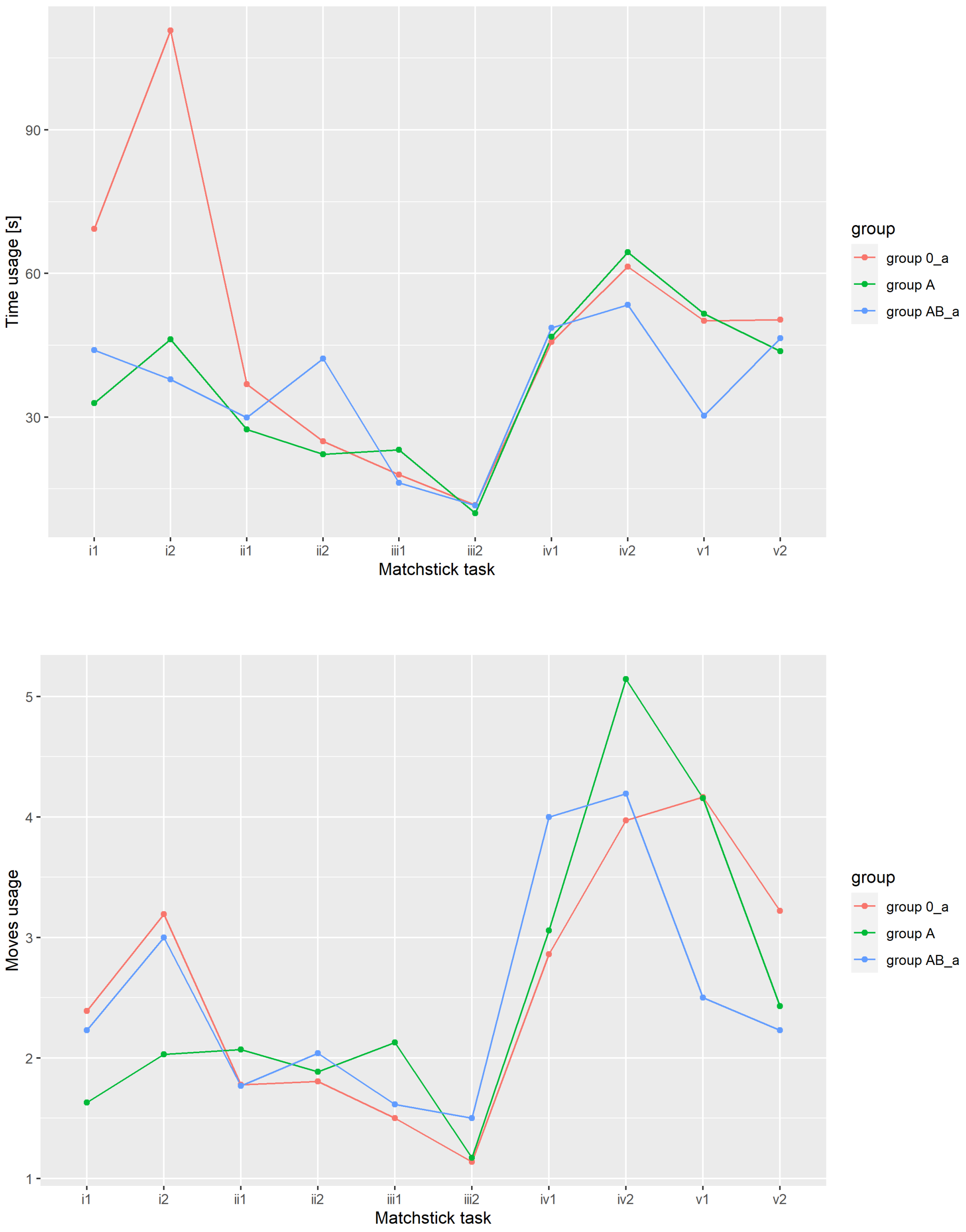


FIGURE 6.6: Performance in groups of strategy 'a'
(time - above, moves - below)

6.2.5 Hypothesis 02.03 - Negative transfer

Our experiment was designed around progressively changing task types from solvable by the initial strategy to solvable by the alternative strategy. While solving the given tasks, the participants, who had learned only one or no strategies, were bound to encounter an impasse when they needed to learn the alternative strategy. The moment and the task, in which this impasse occurred, was referred to as the decomposition point.

The ***decomposition point*** was defined as the number of the task in which the decomposition associated with the alternative strategy happened. In this task, the participant learned the alternative strategy and acquired access to it. This value was calculated as the task where the alternative strategy was used for the first time: the first task in order, where specific_moves_b was not 0 (zero) for group A, or specific_moves_a was not 0 for group B. This value was calculated only for groups A and B.

It is important to keep in mind that the alternative strategy in the decomposition point was used in the process of solving the task, and did not actually aid in this process.

For a better understanding of the decomposition point identification, Table 6.7 shows an example of all the data of one of the participants in group A and the identification of the decomposition point.

In Table 6.7, we have identified the decomposition point in the eighth task. This participant was compared with other participants of group A and their decomposition points in the eighth task against the points of all the participants of group AB in the eighth task.

The participants experienced the decomposition points differently depending on the group they belonged to. Descriptive statistics, showcasing decomposition points for each group, are shown in Table 6.8. The preliminary analysis with the Wilcoxon Signed-Rank test of the occurrences of decomposition points $[V = 66.5, p = .221]$ did not show any significance, but we argue that there was simply not enough data for a proper comparison.

Since only groups A and B had experienced the learning phase and had a more ingrained transfer, we focused only on them in the following analysis of the negative transfer. At this point, we expected the performance of

	Time	**Moves**	**Specific_ moves_a**	**Specific_ moves_b**	
1	18721	1	1	0	
2	19513	1	1	0	
3	20312	2	2	0	
4	7827	1	1	0	
5	9962	2	2	0	
6	4749	1	1	0	
7	46163	2	2	0	
8	79058	3	1	2	→ decomposition point
9	10596	1	0	1	
10	11884	1	0	1	

TABLE 6.7: Example of dependent values
for a participant in group A,
with a marked decomposition point

Task (in order)	group A		group B	
1	0		3	(5%)
2	8	(11%)	2	(4%)
3	4	(6%)	5	(9%)
4	5	(11%)	11	(19%)
5	5	(7%)	4	(7%)
6	5	(7%)	3	(5%)
7	18	(26%)	29	(51%)
8	11	(16%)	0	
9	14	(20%)	0	
10	0		0	
Mean	6.33		5.39	

TABLE 6.8: Descriptive statistics about the occurrence of
decomposition points for each of the groups A and B

groups A and B to be significantly worse compared to the corresponding groups AB_a and AB_b. The decrease in performance was due to a negative transfer, which occurred in the process of overcoming the impasse and included the decomposition of new chunks and the forming of a new mental representation (the definition of what exactly happened at this point varies depending on the different models that we discussed in the section "Theoretical background (3)").

In our analysis, for each task in order, we identified the participants of the experimental group (which had learned only one strategy), who had experienced the decomposition point in this task and compared their performance against all the participants in the control group (which had learned both strategies in the learning phase), who could not encounter a decomposition point during their experiment. This provided us with the results of performance difference between a task with an experienced decomposition point and a task without an experienced decomposition point. From here, we could analyse the effect of the decomposition point or impasse on performance in a single task.

The following working hypotheses separately compare the performance variables (the time and moves variables) for each initial strategy.

1. For every task in order, the participants in group A who experienced the decomposition point spend more time compared to the participants in group AB_a.
   (H_0: For every task in order, the participants in group AB_a and the participants in group A, whose decomposition point occurred in this task, spend an equal amount of time.)
2. For every task in order, the participants in group A who experienced the decomposition point make more moves compared to group AB_a.
   (H_0: For every task in order, the participants in group AB_a and the participants in group A, whose decomposition point occurred in this task, make an equal number of moves.)
3. For every task in order, the participants in group B who experienced the decomposition point spend more time compared to group AB_b.

(H_0: For every task in order, the participants in group AB_b and the participants in group B, whose decomposition point occurred in this task, spend an equal amount of time.)

4. For every task in order, the participants in group B who experienced the decomposition point make more moves compared to group AB_b.
(H_0: For every task in order, the participants in group AB_b and the participants in group B, whose decomposition point occurred in this task, make an equal number of moves.)

Descriptive statistics are shown in the tables of Figures 6.9 and 6.10. The variable *n* describes the number of participants in a comparison, *mean* the mean value, and *sd* the standard definition. The symbol "/" denotes missing data, because no participant has experienced a decomposition point in that task. From the descriptive statistics we can already observe worse performance, indicated by greater mean values in the time and moves variables among participants in the experimental groups in the decomposition task compared to participants in the associated control groups. Similarly, their results were also more diverse, as indicated by the higher standard deviation rates. Additionally, we can notice that the groups with initial strategy 'a' tended to experience a decomposition point in later tasks compared to the groups with initial strategy 'b' – this turned out to be quite a significant difference which is discussed in depth later in this section.

The analysis was conducted with several Wilcoxon Signed-Rank unpaired tests for each of the strategy and variable combinations separately. The results are shown in Table 6.11 of p-values separated according to the strategy and variable used in the comparison. The missing values are again indicated by "/".

During the analysis, we observed reliable significant differences within the 2nd and 4th tasks for the comparison of group A (with decomposition point) against group AB_a in regard to the time and moves variables. In the comparison of group B (with decomposition point) against group AB_b we could also observe significant differences in the 3rd and 7th tasks in the time and moves usage. In the latter comparison, we additionally observed another significance in the time usage in the 2nd task.

group		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A*	n	/	8	4	5	5	5	18	11	14	/
	mean	/	142.76	112.87	101.43	55.08	7.82	45.92	96.94	80.77	/
	sd	/	84.22	54.25	96.31	46.56	3.35	40.32	104.15	79.93	/
AB_a	n	/	26	26	26	26	26	26	26	26	/
	mean	/	37.90	29.94	42.22	16.28	11.52	48.66	53.44	30.32	/
	sd	/	37.11	20.98	107.68	14.84	4.19	44.42	47.49	28.08	/

group		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A*	n	/	8	4	5	5	5	18	11	14	/
	mean	/	7.88	7.00	6.40	4.00	1.40	3.61	6.18	7.71	/
	sd	/	4.55	3.37	2.70	3.74	0.55	3.31	6.10	10.01	/
AB_a	n	/	26	26	26	26	26	26	26	26	/
	mean	/	3.00	1.77	2.03	1.62	1.50	4.00	4.19	2.50	/
	sd	/	5.18	1.99	2.88	2.21	1.17	3.51	4.02	1.98	/

TABLE 6.9: Descriptive statistics of
group A* (group A with decomposition point) ~ AB_a
on the time (above) and moves (below) variable

group		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
B*	n	3	2	5	11	4	3	29	/	/	/
	mean	91.12	468.26	173.98	89.48	28.35	20.14	93.60	/	/	/
	sd	90.47	414.97	83.50	48.20	10.23	13.51	72.33	/	/	/
AB_b	n	29	29	29	29	29	29	29	/	/	/
	mean	35.53	38.90	52.93	58.47	26.77	18.98	25.77	/	/	/
	sd	31.23	31.92	78.84	49.08	20.27	20.94	26.28	/	/	/

group		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
B*	n	3	2	5	11	4	3	29	/	/	/
	mean	13.00	11.50	9.00	8.64	1.00	2.33	5.24	/	/	/
	sd	8.89	2.12	6.20	6.17	0.00	2.31	5.42	/	/	/
AB_b	n	29	29	29	29	29	29	29	/	/	/
	mean	2.34	2.34	2.76	4.17	1.86	1.55	1.59	/	/	/
	sd	1.61	2.38	4.79	3.64	1.88	1.82	1.32	/	/	/

TABLE 6.10: Descriptive statistics of
group B* (group B with decomposition point) ~ AB_b
on the time (above) and moves (below) variable

Task in order	group A ~ time	AB_a moves	group B ~ time	AB_b moves
1	/	/	0.2073	0.0222
2	0.0000 *	0.0004 *	0.0043 *	0.0135
3	0.0020 *	0.0006 *	0.0014 *	0.0056 *
4	0.0023 *	0.0005 *	0.0300	0.0127
5	0.0110	0.0253	0.5052	0.1533
6	0.0547	0.5993	0.5379	0.3845
7	0.5155	0.5185	0.0000 *	0.0009 *
8	0.1078	0.4391	/	/
9	0.0099	0.0452	/	/
10	/	/	/	/
α	0.0063		0.0071	

TABLE 6.11: p-values of Wilcoxon analyses of the experimental group with decomposition point performances against group AB performances
('*' - marks significance)

If we supplement these Wilcoxon test results with the graphs from Figures 6.7 and 6.8, we can observe the effect of decomposition points on performance.

We could expand the comparison with additional data from groups which could be reasonably assumed not to have encountered a decomposition point in a particular task. We could add data from group 0 (which was not expected to have a decomposition point or at least not to have experienced a strong impasse) and data from the experimental group of participants, who did not encounter a decomposition point in a particular task (they might have experienced the decomposition point earlier or later in the experiment). For the following comparison, we joined data from groups 0 and AB (respective of their initial strategy) and compared this new "group" to the experimental groups according to their decomposition points with Wilcoxon Signed-Rank unpaired tests, just like in the previous experiment. The new results are outlined in the graph in Figure 6.9. However, these results seemed to only slightly improve significance in the moves analyses and did not improve significance in the times analyses. The graphs in Figure 6.6, show a relatively close resemblance of these groups to group AB, so the only aspect

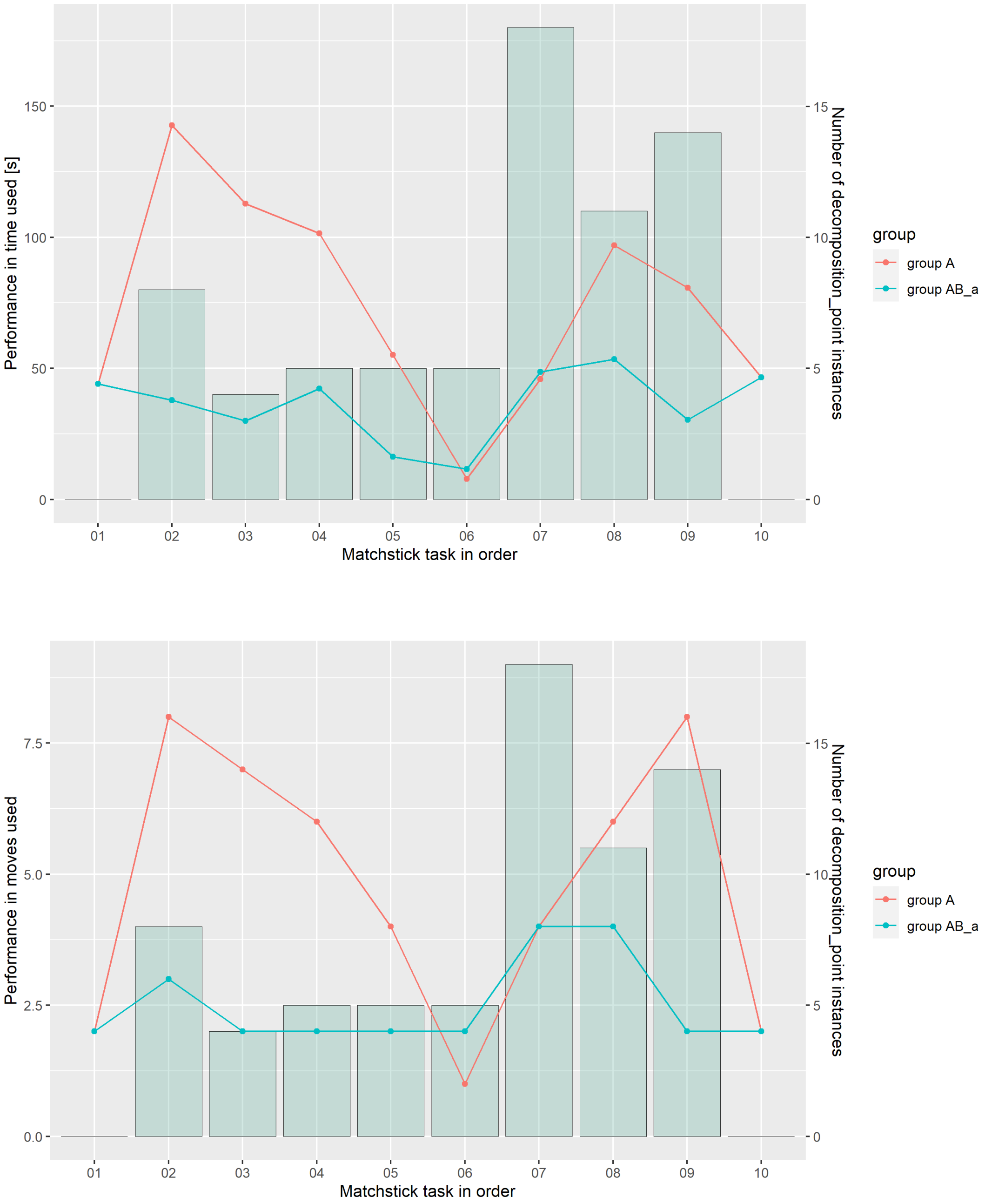


FIGURE 6.7: Performance in decomposition points in group A compared to AB_a
(time - above, moves - below)

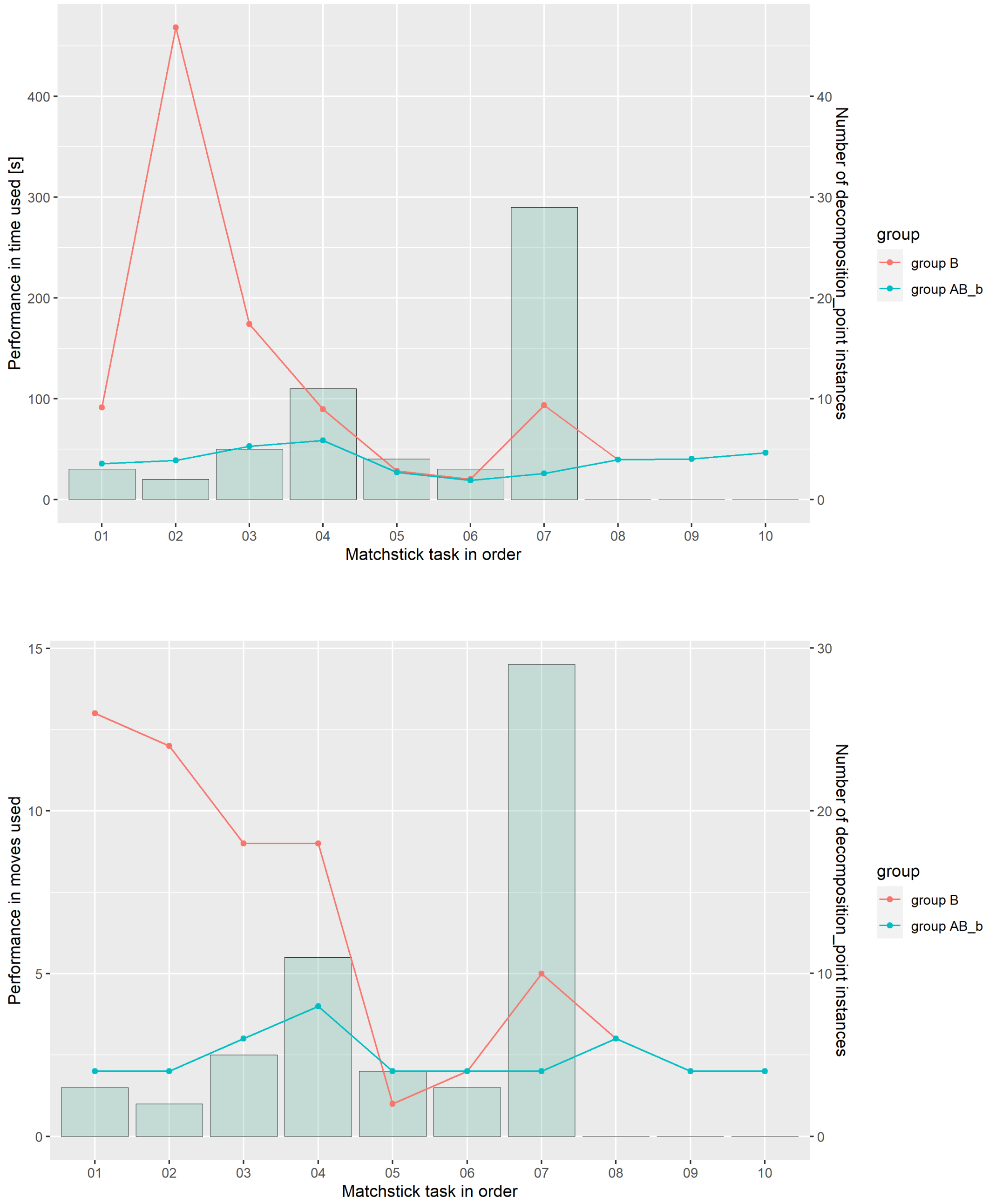


FIGURE 6.8: Performance in decomposition points in group B compared to group AB_b
(time - above, moves - below)

that could improve statistical analysis with the addition of these new groups, is additional participant data – in other words, we just added more participants in the analysis, who had similar learning phases as those in group AB. However, these values were slightly compromised, since only in the case of group AB could we be completely certain that they had not experienced a decomposition point anytime during the experiment. Thus, we did not use the new groups in the actual analysis of this hypothesis.

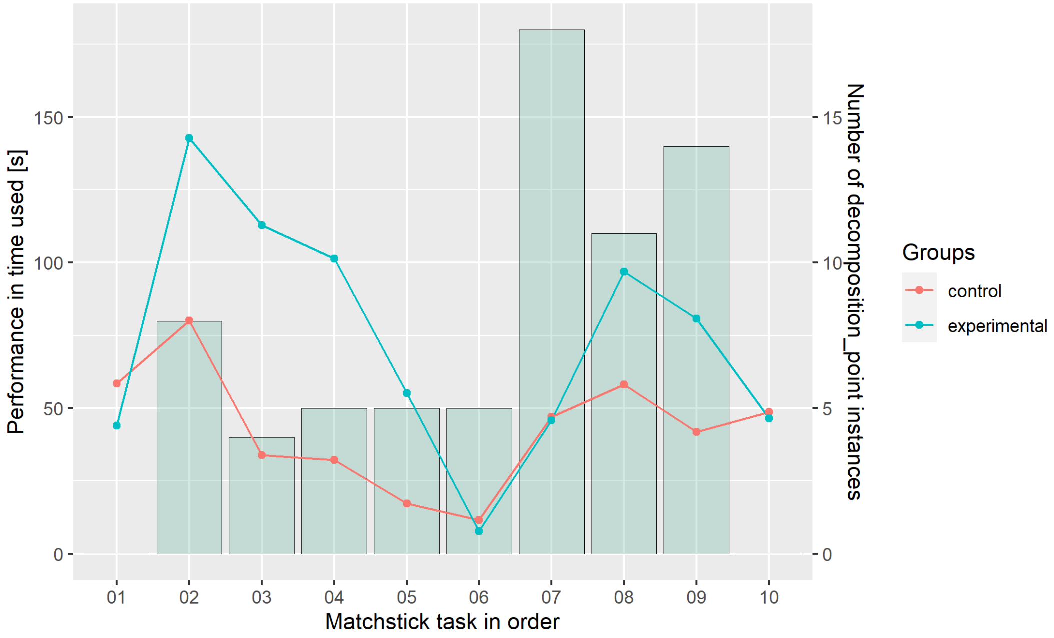


FIGURE 6.9: Performance in decomposition points
in group A ('experimental' group)
compared to the joined groups AB_a and 0_a ('control' group)
(Not used in the actual hypothesis analysis)

The decomposition points were mostly grouped around the 7th task in order (keep in mind that this was a different task for different groups – task iv1 for group A and task ii1 for group B). The main feature of this task was that it was the first task in the experiment which could be more optimally solved with the alternative strategy than with the initial strategy. At this point, every participant in group B, who had not yet experienced the decomposition point, experienced the decomposition point. Surprisingly, this task and the 8th one could still be solved solely by the initial strategy, although not optimally. This was very different compared to group A, whose decomposition points more spread out. Although a lot of participants experienced the decomposition point in the 7$^{\text{th}}$ task, there were a lot of participants who continued using the less-effective initial strategy 'a' in the 7$^{\text{th}}$ and 8$^{\text{th}}$ task as well. These differences are clearly outlined in Table 6.8.

We would argue that these observations supported the assumption of a higher availability of strategy 'a' compared to strategy 'b' (discussed in the analysis section "Research question 04 - Differences in strategies (6.2.7)"). This argument was supported in the observation (1) that participants in group B abandoned their initial strategy 'b' as soon as it became less optimal than the alternative strategy, and in the observation (2) that participants in group A clung longer to their initial strategy, even if it negatively impacted performance.

As for this negative transfer hypothesis, when experiencing the decomposition point the performance was always worse (exhibited in longer time spent and more moves made), compared to participants who had not experienced the decomposition point in this particular task. Some of the analyses did not claim significance, but we assumed that this was due to the low number of instances – with more data, these differences would have been significant (as we have seen with the test addition of group 0).

Based on this, we felt confident to reject the null-hypotheses and confirm our alternative hypothesis.

1. For every task in order, the participants in group A who experienced the decomposition point spend more time compared to group AB_a.
2. For every task in order, the participants in group A who experienced the decomposition point make more moves compared to group AB_a.
3. For every task in order, the participants in group B who experienced the decomposition point spend more time compared to group AB_b.
4. For every task in order, the participants in group B who experienced the decomposition point make more moves compared to group AB_b.

In conclusion, ***experiencing a decomposition point lead to decreased performance in that task.***

6.2.6 Exploratory hypothesis 03 – Priming / Initial learning

> Individuals that learn at least one strategy solve the first task faster than individuals that learn no strategies.

In the following working hypotheses, we compared the times spent on the very first tasks for each of the strategies separately. We could not compare moves, since the first tasks had a mandatory restriction to be solved in a single move.

1. For the first task in order, group 0_a spends more time compared to groups A and AB_a.
   (H_0: For the first task in order, the group pairs 0_a - A and 0_a - AB_a spend an equal amount of time.)
2. For the first task in order, group 0_b makes more moves compared to groups B and AB_b.
   (H_0: For the first task in order, the group pairs 0_b - B and 0_b - AB_b make an equal number of moves.)

The descriptive statistics for time usage are shown in Appendix A - Descriptive statistics in Table A.1. In the analysis with two ANOVAs, one for each strategy, we observed significance only in strategy 'a' [$MATS\ Q_n = 10.147; p = .009$] (strategy 'b' did not experience any significant differences: [$MATS\ Q_n = 4.612; p = .111$]). In the subsequent post-hoc analysis, we realized that this significance can be attributed only to the difference between groups 0_a and A [$MATS\ Q_n = -36410.411;$ 95% $CI = -72686.28, -134.54;$ $p = .049$].

Visually, according to the graph in Figure 6.10, the group 0_a performed (insignificantly) worse than groups A and AB_a. But this was not true for group 0_b. This was insufficient to reject the null hypothesis and we had no expectation that additional data would result in a greater significance.

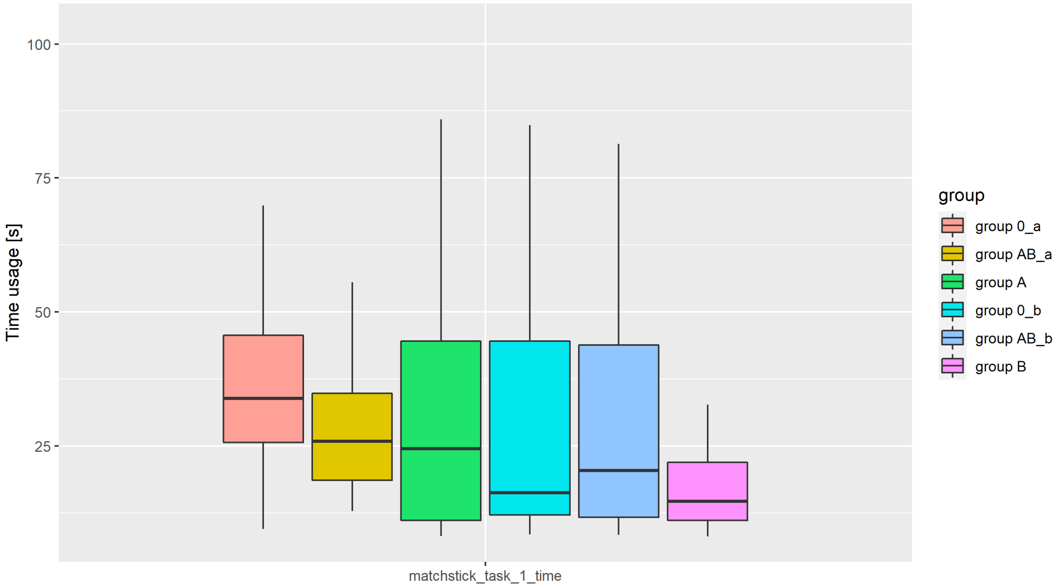


FIGURE 6.10: Time spent on solving the first task (strategy ‘a’ - left of the middle, strategy ‘b’ - right of the middle)

6.2.7 Research question 04 - Differences in strategies

> Is one strategy more available (more commonly used) than the other? And does any strategy lead to a greater transfer than the other?

We started this analysis with a review of Figure 6.11, representing strategy usage by different groups. There we saw a trend in all groups, that participants generally solved the tasks with the strategy that could solve these tasks optimally. This was visible from the graph lines for strategy ’a’ which were steadily falling from task i1 to v2, and vice versa for strategy ’b’. Additionally, we could observe that the groups tended to stick with their initial strategy and solved the given tasks with it, even if it turned out to be less optimal. This was visible in the upper graph, as the participants in groups A, 0_a and AB_a tended to use their initial strategy ’a’ more than the groups B, 0_b and AB_b used the strategy ’b’ in every task. The exact opposite is true for the lower graph.

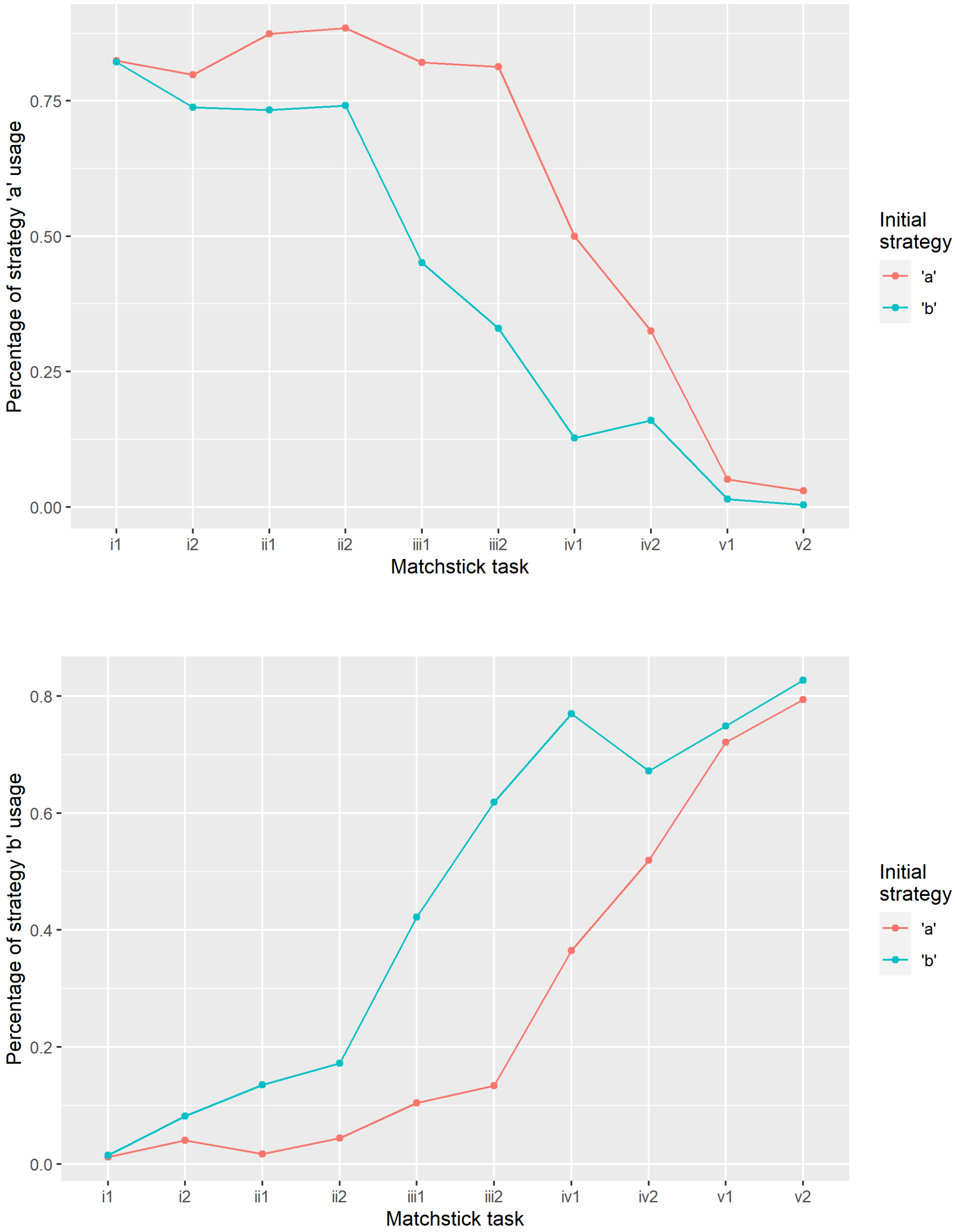


FIGURE 6.11: Strategy usage between groups with different initial strategies (’a’ - groups A, 0_a, AB_a; ’b’ - groups B, 0_b, AB_b)
(strategy ‘a’ - above, strategy ‘b’ - below)

The descriptive statistics for ratio values are shown in Appendix A - Descriptive statistics in Table A.3. In order to find out if there were any significant differences between strategies, we focused on two aspects, each with its separate working hypothesis.

In the first aspect, we compared the usage of one strategy against the other in groups that had learned both strategies, in tasks, that were optimally solvable by both strategies (i.e. tasks iii1 and iii2). If there were no differences between strategies, we expected to see them used equally often in these tasks.

> For each task of task group (iii), the ratio of moves corresponding to strategy 'a' was expected to be equal to the ratio of moves corresponding to strategy 'b' for group AB.

To compare two values of a task, we used the Wilcoxon Signed-Rank test analysis. In task iii1, we observed $[V = 1121.5, p < .001]$ and in task iii2 $[V = 961.5, p = .018]$. For both tasks that correspond to task group (iii), we observed significant differences between the ratio of moves of strategy 'a' and the ratio of moves of strategy 'b'. According to the statistical analysis and the visual analysis of the graph in Figure 6.12, we could already conclude that the strategies do indeed differ, as groups A, 0_a and AB_a had used strategy 'a' significantly more than strategy 'b', while groups B, 0_b and AB_b had used strategy 'b' significantly more than strategy 'a'.

In the second aspect, we compared the usage of initial strategies in the experimental groups in each task in order. If there were no differences between strategies, we should have seen them used equally often as the initial strategies in their respective groups, for each of the tasks in order. Again, we did not compare the first and last two tasks, since they were required to be solved with a single move.

> For each task of task group (ii), (iii) and (iv) in order, the ratio of moves corresponding to strategy 'a' for group A was expected to be equal to the ratio of moves corresponding to strategy 'b' for group B.

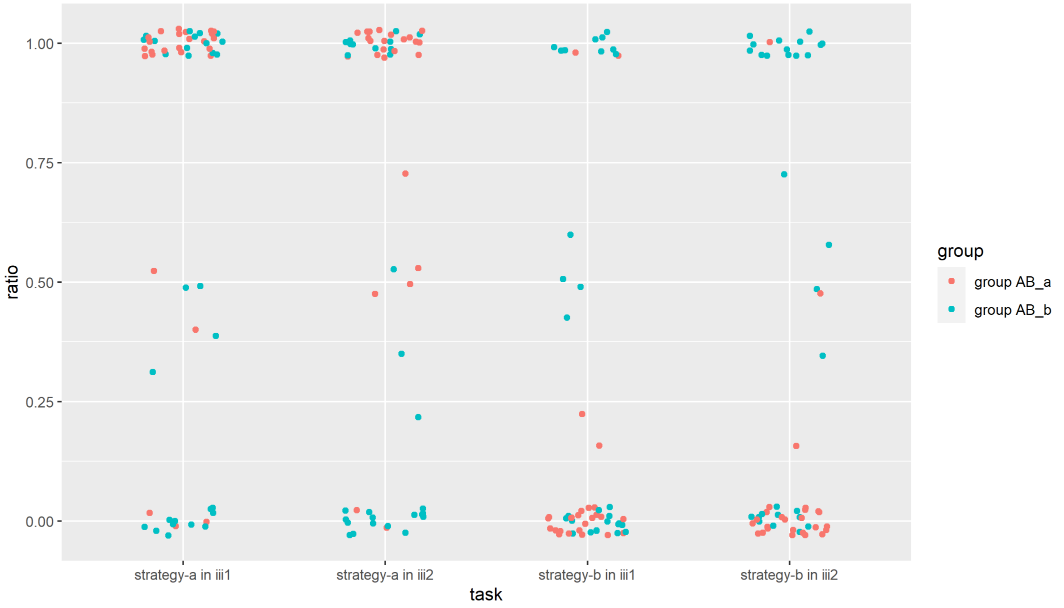


FIGURE 6.12: Differences in strategy ratios for tasks iii1 and iii2 for participants in groups AB_a and AB_b

We analysed this equality by comparing the ratios of initial strategy usage in the experimental groups per tasks in order. We performed Wilcoxon Signed-Rank tests and displayed them in Table 6.12. These results were supported with a visual analysis of the graph in Figure 6.13.

In the results, 3 out of 6 ANOVA measurements indicated significance. According to the graph, a small difference in strategies was observed, mainly that strategy 'a' was used slightly more compared to strategy 'b'. With more measurements, we believe it would become clearer that there was indeed a difference between strategies.

Task	Ratio of initial strategy	
3	$V = 1963.5; p = .845$	
4	$V = 2545.5; p = .002$	*
5	$V = 2404.0; p = .022$	
6	$V = 2231.5; p = .172$	
7	$V = 3015.5; p < .001$	*
8	$V = 2817.5; p < .001$	*

TABLE 6.12: Results of Wilcoxon analyses of ratios of initial strategy usage per task in order in the experimental groups (groups A and B combined) *('*' - marks significance with α = 0.0083)*

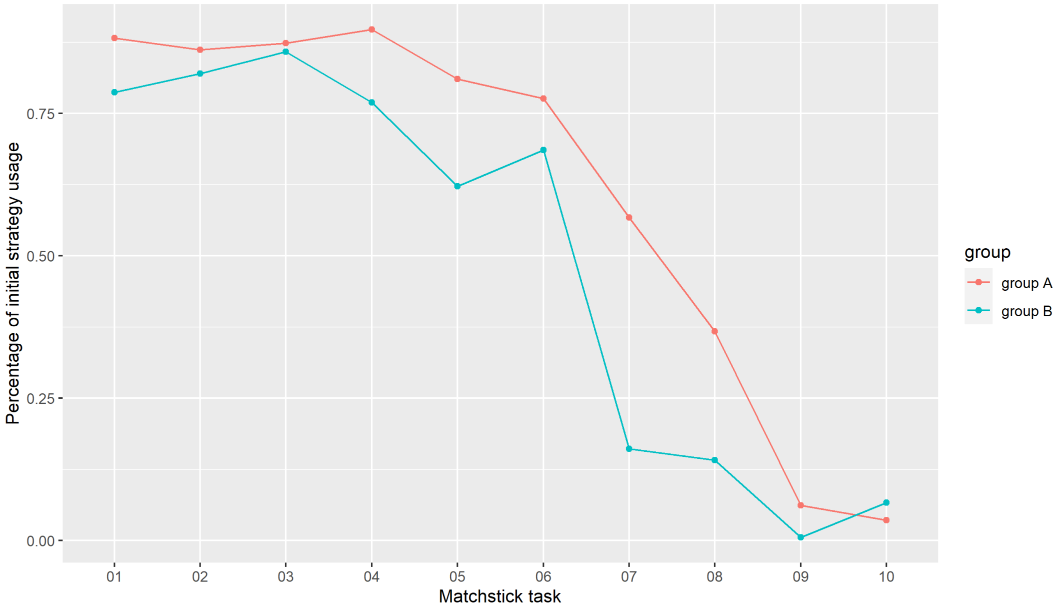


FIGURE 6.13: Initial strategy usage between groups A and B

According to the results, strategy 'a' had higher usage rates than strategy 'b' - this would suggest one of the following:

1. Strategy 'a' was much more available than strategy 'b';
2. Strategy 'a' transferred much better than strategy 'b'.

To answer this question, we needed to compare the same strategy usage ratios in group AB. This group had learned both strategies and should not have experienced strong transfer rates. If there were significant differences in initial strategy usage in group AB, these differences could better be explained by a difference in strategy availability. The results of the ANOVA analysis are shown in Table 6.13, with additional visual support from the graph in Figure 6.14.

Task	**Ratio of initial strategy**	
3	$V = 519.5; p = .003$	*
4	$V = 505.0; p = .019$	
5	$V = 573.5; p < .001$	*
6	$V = 503.5; p = .014$	
7	$V = 558.5; p < .001$	*
8	$V = 526.0; p = .005$	*

TABLE 6.13: Results of Wilcoxon analyses of ratios of initial strategy usage in groups AB_a and AB_b
('' - marks significance with α = 0.0083)*

There was greater significance observed in quite a few of the measurements. Even when comparing groups AB_a and AB_b there seemed to be greater differences between strategies. Additionally, group AB_a seemed to use its initial strategy ('a') relatively more than group AB_b used its initial strategy ('b'). As group AB was not expected to experience any transfer, all differences could only be assigned to strategy availability.

Since strategy availability differed between the two strategies, we could not comment on the difference in transfer between them. The more available strategy also affected transfer, so we could not distinguish the effect of availability and transfer between the strategies. The effect of availability on

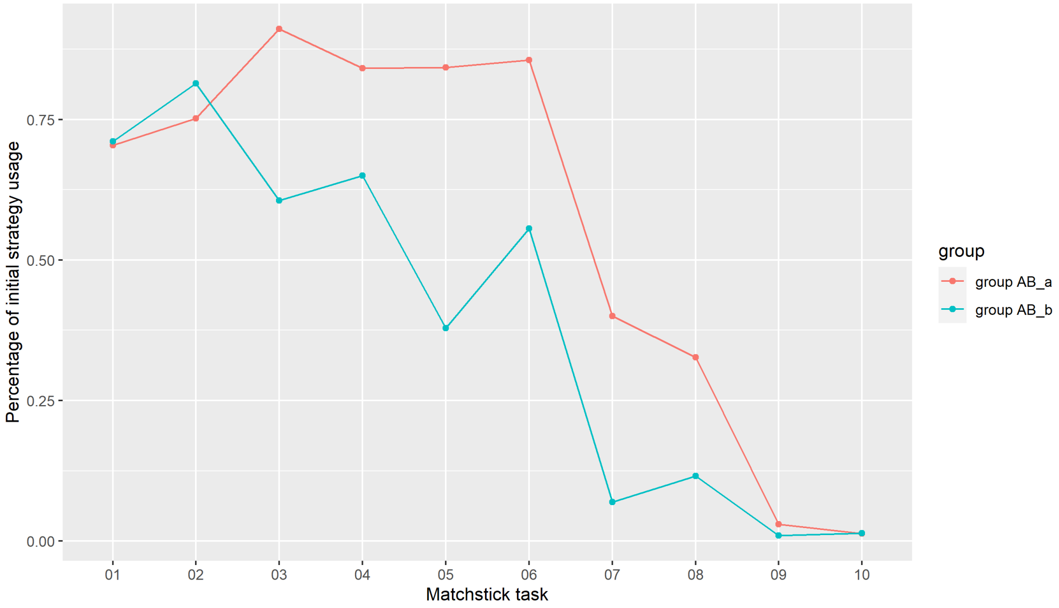


FIGURE 6.14: Initial strategy usage in group AB

transfer could be observed in the following two features of the more available strategies: (1) participants tended to stick with the more available strategy even when the solutions achieved with it were less favourable, and (2) participants tended to switch to the more available strategy at the first opportunity, when their existing strategy was no longer favourable. Keep in mind that "availability" is a comparative measurement of mental representations: We could not describe a strategy as "available", but only as "more or less available than another" strategy.

With this analysis, we could conclude that ***strategy 'a' was more available than strategy 'b'***.

Chapter 7

Discussion

In this chapter, we relate our findings to already existing studies and explain these results in layman's terms. We focus on each significant aspect of our findings separately in order to better facilitate a discussion. At the end of this chapter, we discuss some questions still left open, and provide ideas and guidance for future research in this field.

7.1 Strategy availability

According to the results of the analyses, the given strategies were not as similar as we initially expected. The analysis of Research question 04 - Differences in strategies (6.2.7) strongly suggested that strategy 'a' was more available than strategy 'b'. This means that participants favoured moving matchsticks between or within numerals significantly more than moving matchsticks between or within operator symbols. These results were separately supported with the analysis of Hypothesis 02.03 - Negative transfer (6.2.5), where we have observed significant differences between strategies and assigned these differences to the strategy availability. These results completely contradicted our initial assumption of strategy availability on which we have built our experiment design.

When designing the experiment, we attempted to create strategies with similar requirements and perceived difficulty by implementing the following three equality features: equally wide constraints, equally tight chunks, and a similar amount of possible equation elements. From the start, we based the strategies on the mental representations of a matchstick experiment designed by Knoblich et al. (1999). From their experiment, we initially borrowed the two constraints: numeral constraint and operator constraint. Each of them needed to be relaxed in order to learn their associated strategies, strategy ‘a’ and strategy ‘b’ respectively. Both of these constraints were considered of equal scope, as one could not be considered wider or narrower than the other, so none should have had a higher probability of being relaxed than the other. From Knoblich et al. (1999) we also borrowed chunked elements: a numeral and an operator, respectively associated with strategy ‘a’ and strategy ‘b’. In order to learn a strategy, the respective chunk needed to be decomposed. Both chunks were considered equally tight, as decomposing them would result in elements (single matchsticks) that could not stand alone in a matchstick equation. On top of that, we adapted the matchstick equations themselves to further minimize the differences between strategies. Following Derbentseva (2007), we added multiplication and division symbols to expand the possible set of operators (the experiment by Knoblich et al. (1999) had 3 operators, while ours had 4). On the other hand, with careful equation selection, we had restricted the pool of all commonly used numerals to only 3 (it was still possible to create others, but far less likely). For more on this part of the experiment design, see Appendix C - Technical design.

Nonetheless, we could not fully equate the strategies. They operated on different types of arithmetic symbols: operator and numeral, and these symbols had inherent differences amongst themselves, namely, the number of matchsticks each symbol consisted of (7 matchsticks for a numeral and 4 matchsticks for an operator) and their arithmetic usage (the numeral was a value element and the operator a value-manipulation element).

From our findings, we have concluded that ***the mental representation of strategy ‘a’ was significantly more available than the mental representation of strategy ‘b’*** *(moving matchstick between or within numerals was easier / more likely than moving matchsticks between or within operators)* and we could confidently assume that the reason behind this lay in the remaining differences between these mental representations (outlined as differences between strategies in the previous paragraph). However, we had no reason to blame one particular difference between the strategies to be responsible for the difference in their availabilities. In their work, Kotovsky and Fallside (1989) had also acknowledged differences in the availability of different mental representations. Similarly to our study, they had no grounds to reason what the source of these differences was. Some mental representations were simply more available than others. Discovering the reasons why some mental representations were simply more available than others is a task for future research studies.

From our results we could also observe that this availability difference had an important effect on performance in the tasks and on the transfer itself. This is also supported in one of the findings by Kotovsky and Fallside (1989, p. 32), who reported that "the availability of a representation is a determinant of how readily it will transfer to other problems". By knowing the relative availabilities between mental representations we can select and use the one with the more desired amount of transfer (keep in mind that this is relevant for positive AND negative transfer). This can prove especially useful in future studies where researchers will be able to manipulate the amount of transfer simply by teaching their participants different initial mental representations.

7.2 Inexact positive transfer

In our study, we measured transfer by comparing the experimental group, which was affected by transfer (group A or B), against its respective control group (group AB and 0), which was designed not to experience or be affected by transfer. Our reasoning was that the control groups had not been specifically trained in one particular strategy.

Group AB had been trained in both strategies and decomposed all the required chunks in the learning phase; thus, according to the findings of Knoblich et al. (1999), it had both strategies available and was able to switch between them when required. This control group should have been able to use the most appropriate strategy based on each task individually, experiencing no transfer among the tasks in the experiment. On the other hand, group 0 had skipped the learning phase altogether and learned the initial strategy while solving its 1st task. Afterwards, it should have experienced some degree of strategy transfer, but far less compared to its respective experimental group, since it had not been trained as much. Compared to group 0, the respective experimental group was primed and taught the strategy with two videos; it then practised the strategy in two additional learning tasks. Because of that it was reasonable to assume that the experimental groups should have experienced greater transfer than group 0.

However, our results, especially the analysis of Hypothesis 02.01 - Transfer ratio (6.2.3), showed that our assumption was incorrect and that ***even control groups 0 and AB experienced a significant transfer effect***. Their performance and strategy use strongly correlated with their respective experimental groups, while significantly differing from their counterparts with the other initial strategy. This was completely opposite of what we had originally assumed.

We lack support from any preceding studies to explain these results, as these control groups were introduced into our study without any prior supporting background, as discussed in the section Strategy-learning groups (5.2.1). Nonetheless, we are fairly confident that we are able to explain this phenomenon with *a positive transfer in control groups*.

Introducing the element of transfer into the control group analysis explained the insignificance of the results of the analyses of Hypothesis 01 – Better with two strategies (6.2.1), Hypothesis 02.01 - Transfer ratio (6.2.3) and Hypothesis 02.02 - Positive transfer (6.2.4). Table 6.1 and Figure 6.3 regarding hypothesis 01, Table 6.3 and Figure 6.4 regarding hypothesis 02.01 and Table 6.6 and Figure 6.6 regarding hypothesis 02.02 all show great unexpected similarities between groups 0_a, A & AB_a and between groups 0_b, B & AB_b in the form of insignificant results of statistical analyses and the graph slopes, where the slope of control groups 0 and AB mimics the slope of their related experimental group. These similarities can be completely explained by acknowledging that transfer had some effect on performance in groups 0 and AB. It seems that performing just a few tasks is enough to facilitate the amount of transfer that is statistically indistinguishable from the transfer generated through learning phases.

This explanation also calls into question the following premise, made by Knoblich et al. (1999, p. 1535), which states that "once a problem representation has been changed, the change should persist and so should transfer to all relevant subsequent problems". Group AB was defined by learning both available strategies and creating both mental representations, thus, according to this premise, it should not have experienced any difference in transfer between strategies 'a' and 'b'. For each task in sequence, both strategies should have been readily available and a participant should have picked one of them equally likely as the other. However, this was not the case. Instead, we observed that the participants tended to transfer their initial strategy (with which they solved the first two tasks) to subsequent tasks. This provided clear evidence of positive transfer, directly contradicted the outlined premise, and begged for further investigation why the transfer reoccurred even after both mental representations had been created.

At this point we can only call into question the assumptions about positive transfer and equal usage of both mental representations, however, we cannot say anything about negative transfer. Our investigation of negative transfer is based on the decomposition point, which in its current definition cannot be used to investigate group AB as it does not cover possible subsequent impasses. Two open questions remain: (1) whether negative transfer occurs similarly to the positive transfer and, if answered with "yes", (2) does impasse occur as well, meaning that the problem-solver has to decompose the associated chunks once again? These questions and possible issues remain open and are not addressed in the following section "Negative transfer in decomposition points (7.4)".

This explanation also created an issue within our study, specifically with positive transfer measurement. We had lost the reliable no-transfer performance baseline, to which we would have compared the performance transfer in experimental groups. The transfer measured with this method was the difference in transfer between control and experimental groups. Although there was some visual support in graphs that the experimental groups indeed experienced a greater transfer than the control groups, it turns out that this difference was not significant enough to draw a conclusion.

This issue came with a positive side as well. It had given us an insight into the inner workings of transfer. According to our findings, ***even one successfully solved task can induce a transfer of the used strategy to future tasks***. This was true for both control groups; in their 1st and 2nd tasks, group 0 had to learn a strategy and group AB only had to use one of the already known strategies, and yet they both continued to transfer this strategy to subsequent tasks. Contrary to our expectations, ***a prolonged learning process is not necessary for transfer*** and, according to our results, ***a prolonged learning process does not improve transfer***.

7.3 Unknown priming effect

Combining the explanations provided in the sections “Inexact positive transfer (7.2)” and “Strategy availability (7.1)”, unfortunately still does not explain the results of the analysis of Exploratory hypothesis 03 – Priming / Initial learning (6.2.6), where group 0 did not perform worse than the other groups in the very first task. The introduction of transfer into control groups should have had no effect in the very first task for group 0, as it simply had not yet solved any tasks from which the knowledge could have been transferred. On the other hand, the difference in strategy availability could have explained the poor performance of group AB_b seen in Figure 6.10. This group had already learned the more available strategy ’a’, but had to employ the less available strategy ’b’ in order to solve the first task. This inherent better availability of strategy ’a’ might provide an explanation why group AB_b struggled to employ strategy ’b’. However, even this analysis did not provide significant results, so we will simply have to wait for future studies to provide an answer.

7.4 Negative transfer in decomposition points

Unlike positive transfer, negative transfer created a sufficiently strong effect on performance and was observed in our study. From the results of the analysis of Hypothesis 02.03 - Negative transfer (6.2.5), we observed a significant decrease in performance in both variables: time and moves.

The decomposition point was carefully selected in order to fit the observed representational features associated with an impasse and chunk decomposition. We have defined it based on the premise that participants decompose the new chunks, create new mental representations and learn a new strategy in the same task where the strategy was used for the first time. This was a reasonable assumption, supported by the theory of reproductive thinking, which reasons that creative productive thinking is blocked while reproductive thinking can be successfully applied; the assumption is additionally supported by Luchins (1942), who argued that if only a single mental representation is developed, problem-solvers tend to stick with it and use it in future surface-similar problems. While it was possible that a participant explored other approaches and decomposed the new chunks in previous tasks, before they actually used the new strategy, it was highly unlikely that this happened. Identifying decomposition points by finding the first usage of the non-initial strategy was simply a matter of data analysis. And because of the way our experiment had been designed, it was possible to find a decomposition point for every participant.

A good improvement on this would be to introduce qualitative research methods into the experiment and measure the occurrence of the “AHA” moment – the moment when participants overcome the impasse. Although the point measured by the “AHA” moment should by definition coincide with the given decomposition point, it would be beneficial for the sake of greater clarity of the results.

From our observations, we concluded that ***performance in the decomposition point was significantly decreased***. We were able to successfully explain this performance decrease with the occurrence of an impasse, which by definition blocks the ongoing problem-solving process and requires additional mental work to be resolved. These results coincide with the results of Knoblich et al. (1999) and Alzayat (2011), where negative transfer was observed in carefully crafted impasses (after their participants had solved a set of surface- and deep-similar problems, they gave them a surface-, but not deep-, similar task, which directly induced negative transfer); they confirmed that it indeed negatively correlates with the participant's performance in terms of the times and moves used in the problem-solving process. We have reached the same conclusion, using a slightly different method.

7.4.1 Point of negative transfer occurrence

Unlike studies conducted by Knoblich et al. (1999) and Alzayat (2011), our experiment had the capability to measure at which point the impasse occurred in a set of tasks with an ever-decreasing optimality of the initial strategy.

As we have already discussed in the section "Strategy availability (7.1)", the transfer was dependent on strategy availability. Participants in experimental groups with different initial strategies tended to encounter the point of impasse in different tasks in order.

One possible explanation of this phenomenon is that some tasks are on their own simply much more difficult to solve than others. So, although their initial strategy could still be used to optimally solve the task, some participants might be, for some reason, unable to find a solution to that task with their initial strategy and would thus experience an impasse.

However, since the design of our experiment did not make it possible to evaluate task difficulty, we focus on the second explanation of this phenomenon, which argues that tasks which are less optimally solvable by the initial strategy have a greater probability of an occurrence of impasse. This probability is also affected by the availability of the initial strategy.

A neat side-observation of our experiment is that it is somewhat possible to predict the point of impasse by evaluating the probability of an occurrence of impasse for every task. Firstly, we need to analyse both mental representations, the existing one and the one created by insight in an impasse, and calculate their availability. Secondly, we have to identify surface- and deep-similarities between tasks and how well each mental representation can be used for each task (this was done in detail for every matchstick equation in our experiment). Based on this data, we can sufficiently well predict the probability of a problem-solver switching their strategy for each task.

Our initial prediction was that the decomposition point would probably occur between tasks 7 – 9 in order. In the 7th and 8th task, the initial strategy becomes less optimal, and in the 9th task, completely useless. This prediction was made based on our initial assumption that both strategies are equally available. For strategy 'a', this prediction was fairly accurate. However, we did not expect that strategy 'a' was significantly more available than strategy 'b'. This greater availability of strategy 'a' resulted in the participants in group B being able to switch their initial strategy to the more available strategy much earlier and with less negative transfer. This correlates with much higher rates of decomposition points in the 7th task for group B, as none of the participants continued to use the inferior strategy in the 7th and 8th tasks.

Based on these observations, we argue that ***the availability of a mental representation is positively correlated with negative transfer***. This means that participants with a mental representation of a low availability are more likely to reach an impasse as soon as it starts to perform sub-optimally, and seek to replace it with a different mental representation. This results in lower rates of negative transfer, which was in our experiment observed only in the task with the decomposition point. On the other hand, participants with a mental representation of a high availability are more likely to continue using it, even when it starts to perform sub-optimally. This directly means higher rates of negative transfer, which was in our experiment observable in tasks from 7 to 9 - despite being sub-optimal, some participants stuck with their initial mental representation as long as possible, until the 9^{th} task, where it was completely impossible to use it.

Such observations have not been reported by any other study that we have read and the only reason we were able to observe them lies in our unique experiment design. Nonetheless, our study was very limited (with only 2 different mental representations) and we are eager to read future studies that provide some additional information on the effects of the availability of a mental representation on positive and negative transfer.

7.5 Implications

We sought to build our study in an interdisciplinary manner and in this effort we had combined various research fields in our study design. Considering the area of mental representations, impasses and transfer, this study was mainly based in the field of psychology, therefore we also mostly relied on psychological terms, concepts and practices. The study was expanded with inclusion of the field of problem-solving, which has a lot of subtle connections in almost every cognitive research field. In our case, we made use of its connection to mathematics by deciding to use equation-based problems. Additionally, since our study had been fully carried out online, computer science can be perceived as a supportive meta scientific field.

Our findings are mainly meant to aid further research of mental representations and transfer. The study was fairly focused on a very specific set of matchstick-equation problems; thus the findings can be difficult to generalize to real-world problems, which tend to be much more dynamic and complex. We sincerely hope that future researchers will improve on our methods, fix our shortcomings, and perhaps introduce different types of problems with different mental representations to this research area. The end goal is for our findings and the findings of such future studies to help us to better understand the process of problem-solving in general, opening the possibility for the application of our findings in pedagogy. If we understand how problem-solving works, we can improve these processes in ourselves and teach others as well, so that we can grow as a civilization.

That being said, our current results suggest that ***learning, repetition and training do not result in greater improvements in performance as was expected***. People train to get better and improve themselves. This is true for physical fitness, school subjects, work processes, and various hobbies and activities. In every one of those fields, more time and effort put into practice leads to better performance in that specific field. However, our research suggests that in problem-solving this is not the case. Participants who had been taught a strategy with videos and practised on example tasks did not perform any better than the participants who were simply given a task and expected to solve it. Learning and practising seems to be unneeded and not at all beneficial, so our best course of action seems to be to avoid it. This somewhat coincides with the findings of Knoblich et al. (1999) that once chunks are decomposed, they stay decomposed (here we are looking at their premise only performance-wise and not from the perspective of transfer, which we called into question in the section "Inexact positive transfer (7.2)"), and any additional training in decomposing such chunks makes no difference. All these studies suggest ***that training problem-solving does not improve performance***. Nonetheless, future studies are needed to see if this is the case only for matchstick-equation problems or if it is relevant for other problems as well.

Our findings also suggest that ***individuals tend to stick with their existing and often used behaviours and practices***. People do not like to change their behaviour and prefer to utilize the less cognitively demanding reproductive thinking (Wertheimer and Wertheimer, 1959) in as many situations as possible; for example, while driving to work, travellers often take the same route every time, even if some other route might have less traffic at certain times.

Not only that, our findings suggest that ***even individuals who possess the knowledge of two or more ways of solving a certain problem are prone to transfer***. Increased knowledge of a certain subject is considered to decrease biases and enable people to solve problems more flexibly and objectively. However, our study suggests that this might not be the case. According to the section "Inexact positive transfer (7.2)", even participants who had learned all of the strategies, tended to pick one and stick with it. People seem to remain biased no matter how much knowledge they collect. And this is not too difficult to imagine; for example, every teacher has their favourite subject about which they are more passionate than others, and every student and writer have their favourite pen or keyboard layout that they, for some reason, simply prefer over others. As stated before, we have no supporting studies to support these findings and we eagerly await what future studies will have to say about them.

Furthermore, our findings suggest that ***the most available mental representations lead to the most costly pitfalls***. Mental representations with higher availability are more common and easier to obtain according to Kotovsky and Fallside (1989), and also cause more severe cases of negative transfer according to our study. In general terms, following mainstream practices and procedures is easy, but once these practices are no longer useful it is much harder to break from them and find more applicable ones, than it is to break from more obscure practices and procedures in similar situations. Well, in real-world situations with multiple people, social biases, such as the conformity bias and authority bias, have a strong effect on such mainstream or obscure practices and on the availability of mental representations related to them. But our study suggests that even without these social factors it is more difficult to overcome an impasse on a more available mental representation.

Lastly, we look into the online aspect of our study. We are far from the first to perform a fully online study, but we can nevertheless share our experience with it. Performing a fully online study is extremely difficult and should be carefully thought through. There is a huge amount of work to set up a website and create an application, restructure the study design to support the online aspect, and, lastly, consider all the possible interactions that could be made by a participant (in the scope of the study or with possible malicious intent). Our study took years longer than we had initially assumed, simply because of complications encountered with the online aspect. On the other hand, online studies are widely available and can test far more participants than a study conducted in person could ever do. Our study was conducted in early 2020 during the Covid-19 pandemic, when we could not interact with participants in person for health reasons - we would not have been able to conduct the study without the online aspect. Additionally, online studies are fully experimenter independent, meaning that the experimenter has no possible influence on the results and that every instance of the experiment is performed in exactly the same way with no variations (this can prove to be problematic as well, since it is not possible for an online program to adapt to any unforeseen situations). Overall, we advocate in favour of online studies, but warn future researchers to carefully consider if the online aspect is really beneficial for their study.

7.6 Improvements

Several shortcomings of our study have been identified during the course of the experiment. Participants have outlined them in the comment section or reported them verbally directly to us. All of those shortcomings and their possible solutions are outlined in "Appendix E - Possible improvements", while in this section we discuss just the most crucial ones.

For future studies, we encourage researchers to provide more ways to gracefully exit the experiment or provide some hints, even though that would completely invalidate the results. Our approach to keep participants in the experiment until they find a solution has proved to induce a lot of frustration. Measuring the performance decline during an impasse is a huge part of our study and we had to make participants overcome it on their own in order to obtain valid results. Therefore, the participants were left to solve their tasks without any support. Our expectation was that the participants would just keep working on the experiment until they found a solution. And since the experiment is void if the participants do not finish it or if they receive any additional help, we should not encourage this in any way. Unfortunately, some participants were not able to overcome that impasse, even though there was no time limit and they could move virtually any number of matchsticks. It turns out that these participants got very frustrated. If we had provided a hint or a graceful exit from the experiment – in both cases, the collected data would not be valid any more – they would have left our experiment with a more favourable experience, making them more likely to encourage others to participate in our experiment.

7.7 Future directions

Many suggestions for future studies were raised in this chapter, especially when we observed unexpected strategy differences and unexpected transfer in control groups.

Further studies could be conducted using different sets of problems and different mental representations. Perhaps more than just two mental representations as in this study. Collecting such data would be very beneficial for our understanding of the inner workings of mental representations and could lead to the design and creation of more focused studies.

An interesting study would be to determine how participants select which mental representation to use. When one has created multiple mental representations, they can use any one of them in order to solve a future problem, according to Knoblich et al. (1999), but they will probably pick only one and stick with it, according to our study. It could provide great insight into mental representations if we figure out how the selection process for mental representations works in cases where no transfer is possible (e.g. the very first task) and a participant has multiple mental representations available.

Another interesting future research study would be the transfer of decomposed chunks between very different types of problems. Ericsson and Lehmann (1996) have studied experts and their study/training practices along various topics, which are more or less related to the problems at hand. Their research suggests that expert problem-solving performance is associated with a large repertoire of already decomposed problem-relevant chunks.

Yet another interesting study would be to determine how long mental representations exist. After a certain time period of not actively using these representations, they tend to slip from the mind and problem-solvers need to recreate them anew. How long does it take and is it easier the second time around? In our study, we did not bother with the long-term life-cycle of these representations. As our experiment took approximately 20 minutes and we required complete focus from our participants, we worked on the assumption that none of them would "forget" the mental representation in such a short time interval. However, we cannot be certain of it.

In our study, we struggled because we had underestimated transfer generation. It would be beneficial for future studies like ours to discover how to contain the transfer – what the minimal requirements are to generate transfer in one group and how to completely prevent it in another. Such findings would prevent pitfalls similar to the one into which our study has fallen.

All of these future studies (and many others) would lead to a better understanding of mental representations and with it we could, hopefully, tackle our main goal that we described in the chapter "Motivation (2)": How to completely avoid perception pitfalls, such as an impasse, in our everyday problems?

Chapter 8

Conclusion

> *"All life is problem solving."*
>
> *– Karl Popper*

Problems, big and small, are omnipresent in our daily lives. We encounter and tackle them sometimes with a clear and rational thought, and sometimes with an automatic well-trained response. Our study has acquired knowledge from previous studies on various problems of different types and focused on the aspects of the process of solving them, with a special focus on mental representations. The study was a success. Although we received some results that did not match our assumptions and predictions, we have made progress in various research fields and received valuable insight into the role of mental representations in problem-solving.

Hopefully, our study has brought new findings and provided grounds for further interdisciplinary research into the nature of mental representations, learned strategies, and negative transfer. And finally, we encourage researchers to build on our study and dig deeper into the fascinating research world of mental representations and transfer. We sincerely hope that our findings about transfer will transfer.

Bibliography

Alzayat, Ayman (2011). "Positive and Negative Analogical Transfer in Problem Solving". MA thesis. University of Waterloo.

Birch, Herbert G (1945). "The relation of previous experience to insightful problem-solving." In: *Journal of Comparative Psychology* 38.6, p. 367.

Blessing, Stephen and Patrick Dronek (2006). "The Effect of Einstellung on Compositional Processes". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 28. 28.

Blessing, Stephen B and John R Anderson (1996). "How people learn to skip steps." In: *Journal of experimental psychology: learning, memory, and cognition* 22.3, p. 576.

Camerer, Colin F and Robin M Hogarth (1999). "The effects of financial incentives in experiments: A review and capital-labor-production framework". In: *Journal of risk and uncertainty* 19.1-3, pp. 7–42.

Clark, E. H. (1929). "Evolution of the Call-Indicator System". In: *Bell Laboratories Record* 8.5, 171–173.

Davidson, Janet E, Robert J Sternberg, and Robert Jeffrey Sternberg (2003). *The psychology of problem solving*. Cambridge university press.

Derbentseva, Natalia (2007). "The intensity of the insight experience in problem solving: Structural and dynamic properties". In:

Dominowski, Roger L (1995). "Productive Problem Solving". In: *The creative cognition approach* 63.

Duncker, Karl and Lynne S Lees (1945). "On problem-solving." In: *Psychological monographs* 58.5, p. i.

Ericsson, K Anders and Andreas C Lehmann (1996). “Expert and exceptional performance: Evidence of maximal adaptation to task constraints”. In: *Annual review of psychology* 47.1, pp. 273–305.

Friedrich, Sarah, Frank Konietschke, and Markus Pauly (2018). “Analysis of multivariate data and repeated measures designs with the R package MANOVA. RM”. In: *arXiv preprint arXiv:1801.08002*.

Gentner, Dedre, Russell Landers, et al. (1985). “Analogical reminding: A good match is hard to find”. In: *Proceedings of the international conference on systems, man and cybernetics*, pp. 607–613.

Gentner, Dedre and Cecile Toupin (1986). “Systematicity and surface similarity in the development of analogy”. In: *Cognitive science* 10.3, pp. 277–300.

Gick, Mary L and Keith J Holyoak (1983). “Schema induction and analogical transfer”. In:

Glucksberg, Sam (1962). “The influence of strength of drive on functional fixedness and perceptual recognition.” In: *Journal of experimental psychology* 63.1, p. 36.

Grant, Elizabeth R and Michael J Spivey (2003). “Eye movements and problem solving: Guiding attention guides thought”. In: *Psychological Science* 14.5, pp. 462–466.

Guthrie, Edwin Ray and George Plant Horton (1946). “Cats in a puzzle box.” In:

Hesse, Mary (2000). “Models and analogies”. In: *A Companion to the Philosophy of Science: Malden, MA, Blackwell Publication*, pp. 299–307.

Holland, John H et al. (1989). *Induction: Processes of inference, learning, and discovery*. MIT press.

Holyoak, Keith J (1985). “The pragmatics of analogical transfer”. In: *The psychology of learning and motivation* 19, pp. 59–87.

— (1990). “Problem solving”. In: *Thinking: An invitation to cognitive science* 3, pp. 267–296.

Holyoak, Keith J and Kyunghee Koh (1987). “Surface and structural similarity in analogical transfer”. In: *Memory & cognition* 15.4, pp. 332–340.

Kaplan, Craig A and Herbert A Simon (1990). "In search of insight". In: *Cognitive psychology* 22.3, pp. 374–419.

Katona, George (1940). "Organizing and memorizing: studies in the psychology of learning and teaching." In:

Kershaw, Trina C and Stellan Ohlsson (2004). "Multiple causes of difficulty in insight: the case of the nine-dot problem." In: *Journal of experimental psychology: learning, memory, and cognition* 30.1, p. 3.

Knoblich, Günther et al. (1999). "Constraint relaxation and chunk decomposition in insight problem solving." In: *Journal of Experimental Psychology: Learning, memory, and cognition* 25.6, p. 1534.

Köhler, W (1925). "The mentality of apes [Translated by E. Winter from the 2nd revised edition, 1927]". In: *New York: Liveright*.

Kotovsky, Kenneth and David Fallside (1989). "Representation and transfer in problem solving". In: *Complex information processing*. Psychology Press, pp. 89–128.

Kotovsky, Kenneth, John R Hayes, and Herbert A Simon (1985). "Why are some problems hard? Evidence from Tower of Hanoi". In: *Cognitive psychology* 17.2, pp. 248–294.

Luchins, Abraham S (1942). "Mechanization in problem solving: The effect of Einstellung." In: *Psychological monographs* 54.6, p. i.

Luchins, Abraham S and Edith H Luchins (1950). "New experimental attempts at preventing mechanization in problem solving". In: *The Journal of General Psychology* 42.2, pp. 279–297.

MacGregor, James N, Thomas C Ormerod, and Edward P Chronicle (2001). "Information processing and insight: a process model of performance on the nine-dot and related problems." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1, p. 176.

Maier, Norman RF (1931). "Reasoning in humans. II. The solution of a problem and its appearance in consciousness." In: *Journal of comparative Psychology* 12.2, p. 181.

Mayer, Richard E (1995). "The search for insight: Grappling with Gestalt psychology's unanswered questions." In:

Mayer, Richard E and James G Greeno (1972). "Structural differences between outcomes produced by different instructional methods." In: *Journal of educational psychology* 63.2, p. 165.

McGraw, Kenneth O and John C McCullers (1979). "Evidence of a detrimental effect of extrinsic incentives on breaking a mental set". In: *Journal of Experimental Social Psychology* 15.3, pp. 285–294.

Merriam-Webster (2020a). *Insight*. URL: `https://www.merriam-webster.com/dictionary/insight`.

— (2020b). *Problem*. URL: `https://www.merriam-webster.com/dictionary/problem`.

Metcalfe, Janet and David Wiebe (1987). "Intuition in insight and noninsight problem solving". In: *Memory & cognition* 15.3, pp. 238–246.

Newell, Allen, Herbert Alexander Simon, et al. (1972). *Human problem solving*. Vol. 104. 9. Prentice-Hall Englewood Cliffs, NJ.

Ohlsson, Stellan (1984). "Restructuring revisited: II. An information processing theory of restructuring and insight". In: *Scandinavian journal of psychology* 25.2, pp. 117–129.

— (1992). "Information-processing explanations of insight and related phenomena". In: *Advances in the psychology of thinking* 1, pp. 1–44.

Randall, Robert L, J Arthur Woodward, and Douglas G Bonett (1997). "A wald test for the multivariate analysis of variance: small sample critical values". In: *Communications in Statistics-Simulation and Computation* 26.4, pp. 1275–1299.

Reed, Stephen K, George W Ernst, and Ranan Banerji (1974). "The role of analogy in transfer between similar problem states". In: *Cognitive psychology* 6.3, pp. 436–450.

Ross, Brian H (1984). "Remindings and their effects in learning a cognitive skill". In: *Cognitive psychology* 16.3, pp. 371–416.

Saugstad, Per and Kjell Raaheim (1960). "Problem-solving, past experience and availability of functions". In: *British Journal of psychology* 51.2, pp. 97–104.

Scheerer, Martin (1963). “Problem-solving”. In: *Scientific American* 208.4, pp. 118–131.

Schooler, Jonathan W, Marte Fallshore, and Stephen M Fiore (1995). “Epilogue: Putting insight into perspective”. In: *The nature of insight*, pp. 559–587.

Strapparava, Carlo and Gözde Özbal (2010). “The color of emotions in texts”. In: *Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon*, pp. 28–32.

Team, R Core (2017). “R: a language and environment for statistical computing. R Found”. In: *Stat. Comput. Vienna, Austria*.

Thorndike, Edward L (1898). “Animal intelligence: an experimental study of the associative processes in animals.” In: *The Psychological Review: Monograph Supplements* 2.4, p. i.

UNESCO (2012). *International Standard Classification of Education - ISCED (2011)*. URL: `http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf`.

Vosniadou, Stella (1989). “Analogical reasoning as a mechanism in knowledge acquisition: A developmental perspective”. In: *Similarity and analogical reasoning*, pp. 413–437.

Weisberg, Robert W (1995). “Prolegomena to theories of insight in problem solving: A taxonomy of problems”. In: *The nature of insight*, pp. 157–196.

Wertheimer, Max and Michael Wertheimer (1959). *Productive thinking*. Harper New York.

Wieth, Mareike and Bruce D Burns (2006). “Incentives improve performance on both incremental and insight problem solving”. In: *The Quarterly Journal of Experimental Psychology* 59.8, pp. 1378–1394.

Appendix A

Descriptive statistics

group	n		1	2	3	4	5	6	7	8	9	10
			i1	**i2**	**ii1**	**ii2**	**iii1**	**iii2**	**iv1**	**iv2**	**v1**	**v2**
0_a	36	mean	69.33	110.72	36.84	24.97	17.94	11.60	45.71	61.44	50.09	50.31
		sd	65.38	125.10	38.26	23.15	20.96	6.04	38.72	62.33	61.36	50.88
AB_a	26	mean	44.03	37.90	29.94	42.22	16.28	11.52	48.66	53.44	30.32	46.51
		sd	46.07	37.11	20.98	107.68	14.84	4.19	44.42	47.49	28.08	45.86
A	70	mean	32.92	46.28	27.42	22.20	23.17	9.95	46.75	64.40	51.56	43.75
		sd	35.09	58.70	32.49	33.99	29.25	4.58	51.94	56.45	63.09	48.38
			v1	**v2**	**iv1**	**iv2**	**iii1**	**iii2**	**ii1**	**ii2**	**i1**	**i2**
0_b	36	mean	56.60	47.23	35.57	96.98	53.84	15.32	24.58	47.64	39.37	83.04
		sd	123.34	84.13	39.12	75.36	58.52	15.65	17.44	41.53	36.28	134.15
AB_b	29	mean	35.53	38.90	52.93	58.47	26.77	18.98	25.77	39.43	40.07	46.36
		sd	31.24	31.92	78.84	49.08	20.27	20.94	26.28	44.16	46.23	118.64
B	57	mean	23.88	51.85	38.74	46.27	25.50	23.45	58.39	49.88	53.93	53.08
		sd	30.86	113.80	56.07	39.82	25.92	39.17	63.94	63.27	106.04	58.05

			i1	i2	ii1	ii2	iii1	iii2	iv1	iv2	v1	v2
0	72	mean	54.35	96.88	30.71	36.30	35.89	13.46	40.64	79.21	53.35	48.77
		sd	54.63	129.54	30.16	35.28	47.24	11.93	38.98	70.96	96.78	69.05
AB	55	mean	41.94	42.36	27.74	40.75	21.81	15.45	50.91	56.10	33.06	42.50
		sd	45.77	89.18	23.80	79.88	18.52	15.80	64.35	47.95	29.63	38.95

TABLE A.1: Descriptive statistics - time variable

group	**n**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
			i1	**i2**	**ii1**	**ii2**	**iii1**	**iii2**	**iv1**	**iv2**	**v1**	**v2**
0_a	36	mean	2.39	3.19	1.78	1.81	1.50	1.14	2.86	3.97	4.17	3.22
		sd	2.26	3.65	1.73	2.16	1.95	0.59	3.10	3.32	6.21	5.24
AB_a	26	mean	2.23	3.00	1.77	2.04	1.62	1.50	4.00	4.19	2.50	2.23
		sd	1.63	5.18	1.99	2.88	2.21	1.17	3.51	4.02	1.98	1.97
A	70	mean	1.63	2.03	2.07	1.89	2.13	1.17	3.06	5.14	4.16	2.43
		sd	1.51	2.72	2.45	1.74	2.24	0.42	2.70	4.61	6.19	2.84
			v1	**v2**	**iv1**	**iv2**	**iii1**	**iii2**	**ii1**	**ii2**	**i1**	**i2**
0_b	36	mean	2.89	2.47	2.00	6.19	4.14	1.44	2.31	4.19	3.36	4.61
		sd	4.31	5.72	1.94	3.75	6.44	1.38	2.04	5.16	4.10	3.40
AB_b	29	mean	2.34	2.34	2.76	4.17	1.86	1.55	1.59	2.69	2.34	1.69
		sd	3.58	1.93	1.32	3.56	1.88	1.82	4.79	4.05	3.50	3.19
B	57	mean	2.51	2.54	2.49	3.75	2.25	1.86	3.63	2.96	2.25	2.81
		sd	2.54	4.01	4.29	5.74	2.17	3.07	3.45	3.64	1.61	2.38

			i1	**i2**	**ii1**	**ii2**	**iii1**	**iii2**	**iv1**	**iv2**	**v1**	**v2**
0	72	mean	2.88	3.90	2.04	3.00	2.82	1.29	2.43	5.08	3.53	2.85
		sd	3.45	4.82	1.84	3.27	4.91	1.07	2.64	4.45	5.26	4.40
AB	55	mean	2.29	2.31	1.67	2.38	1.75	1.53	3.35	4.18	2.42	2.29
		sd	2.81	3.84	1.66	3.24	2.03	1.54	4.24	3.79	1.78	2.17

TABLE A.2: Descriptive statistics - moves variable

group	**strategy**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
			i1	**i2**	**ii1**	**ii2**	**iii1**	**iii2**	**iv1**	**iv2**	**v1**	**v2**
0_a	'a'	mean	.80	.71	.85	.89	.83	.85	.44	.24	.05	.03
		sd	.29	.33	.25	.25	.37	.35	.47	.27	.11	.08
	'b'	mean	.02	.06	.02	.06	.14	.14	.45	.64	.74	.75
		sd	.06	.12	.12	.19	.33	.35	.47	.35	.35	.30
AB_a	'a'	mean	.70	.75	.91	.84	.84	.86	.40	.33	.03	.01
		sd	.29	.34	.25	.30	.34	.30	.39	.33	.09	.07
	'b'	mean	.03	.03	.02	.06	.09	.06	.39	.53	.74	.79
		sd	.09	.07	.08	.15	.27	.22	.43	.38	.27	.27
A	'a'	mean	.88	.86	.87	.90	.81	.78	.57	.37	.06	.04
		sd	.24	.28	.23	.22	.33	.38	.44	.34	.12	.11
	'b'	mean	.00	.03	.02	.03	.09	.16	.31	.45	.70	.82
		sd	.00	.09	.06	.12	.27	.37	.42	.40	.35	.29
			v1	**v2**	**iv1**	**iv2**	**iii1**	**iii2**	**ii1**	**ii2**	**i1**	**i2**
0_b	'a'	mean	.01	.00	.16	.29	.70	.39	.76	.61	.74	.60
		sd	.05	.01	.35	.29	.37	.47	.28	.37	.30	.39
	'b'	mean	.72	.85	.76	.54	.14	.57	.15	.27	.03	.16
		sd	.32	.25	.40	.34	.32	.48	.23	.31	.09	.17
AB_b	'a'	mean	.02	.00	.25	.19	.54	.41	.87	.82	.84	.90
		sd	.07	.02	.38	.24	.47	.48	.22	.33	.25	.24
	'b'	mean	.71	.81	.61	.65	.38	.56	.07	.12	.01	.01
		sd	.28	.26	.44	.31	.46	.47	.16	.24	.04	.05
B	'a'	mean	.01	.01	.05	.06	.25	.25	.64	.78	.87	.74
		sd	.05	.03	.17	.14	.40	.42	.35	.31	.24	.31
	'b'	mean	.79	.82	.86	.77	.62	.69	.16	.14	.01	.07
		sd	.30	.25	.28	.27	.45	.43	.24	.25	.04	.13

TABLE A.3: Descriptive statistics - ratio variable

Appendix B

Sensitive information

B.1 Participant's agreement

According to EU regulations, no approval from an ethics committee is required. It is sufficient to inform participants about what personal data is being collected and what this data is used for. It is also necessary to inform them of potential dangers and provide them with a clear option to opt-out of the experiment in case they do not agree with the terms. In our experiment, there were no foreseen physical dangers for participants and we did not expect any psychological exhaustiveness as well, due to the experiment's simplicity. There was also no perceived danger to the participants' (mental or physical) health. Participation was also effectively anonymous – some personal data needed to be collected for the purpose of the experiment and to prevent fraud, but this data was partially encrypted and not sufficient to determine the identity of participants. Raw data was also kept securely and exclusively available to the experimenters. To participate, every participant needed to agree to the terms about collecting this data.

B.2 Participants' private information

Before the actual start of the experiment, additional personal information about the participants was gathered. This information consisted of their

IP address and associated network information (automatically gathered), and their sex, age, country and level of education. The latter information was freely provided and we simply had to trust in its validity. All of this information was used in the meta-analysis of the participants or in the tracking of said participants and had no connection with the actual experiment data.

We experienced huge problems with selecting the Education level. The level-selection options were based on the UNESCO (2012) ISCED standard with added degree titles. The problematic part was that education systems differ wildly between countries and the same degree titles and levels cannot easily be applied to every one of them. We did our best to provide a coherent grouping for as many education milestones as possible, but were ultimately unsuccessful, as a lot of participants complained about this design. In retrospective, a much better design for collecting relative education levels would be to query participants about the number of years spent in education (excluding kindergarten).

Appendix C

Technical design

C.1 Equation

In the scope of our study, a “matchstick equation” was a problem or a puzzle, consisting of a visual and an arithmetic dimension. Its visual dimension consisted of an organized set of virtual matchsticks (in the following text: “matchsticks”) which had shaped out separate parts or elements of the equation in a standardly recognizable fashion. Each of these elements had an associated corresponding arithmetic dimension, either a numerical value or an arithmetic operation. When these arithmetic components were put together, they formed a verifiable statement, called an equation.

An equation was marked as *“valid”*, if each element in it was a valid element of its type.
An equation was marked as *“solved”*, if it was valid and after resolving all of the operators according to standard arithmetic practices, the left-side and right-side values of the comparator matched.

C.2 Parts of an equation

Each matchstick equation (in the following text: equation) was built of elements, each belonging to one of the following groups:

1) Numerals or digits;
2) Operators or sign symbols;
3) Comparators.

C.3 Symbols

There were 15 symbols available in the matchstick tasks: 10 numerals, 4 operators and 1 comparator. They were each built with a set of matchsticks (from 1 to 7) in a specific commonly recognized pattern.

For consistency all matchsticks were aligned with the matchstick's head directed upwards (primary) and left (secondary).

C.4 Numerals

Numerals (or digits) were core-value elements. Each numeral represented a distinct arithmetic integer value between 0 and 9, as shown in Figure C.2. Each numeral also had a visual depiction with virtual matchsticks, oriented to present a corresponding value in the standard seven-segment display notation (shown in Figure C.1 and discussed by Clark (1929)), which is commonly used in electronic devices that display numerical information.

FIGURE C.1: Numeral frame
also known as the seven-segment display

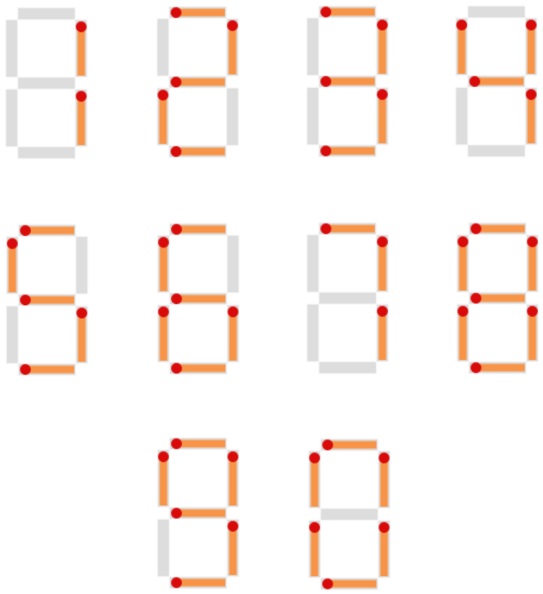


FIGURE C.2: All valid numeral representations

Matchsticks could be scattered in $2^7 = 128$ different possibilities within the seven-segment display notation, but only the 10 variations shown in Figure C.2 represented arithmetic values and were considered “valid”.

In the instructions for the experiment, we referred to these elements as “numerals”, although we could have also used the synonym “digit”. The decision was made to use the term “numeral”, based on it being a foreign word in selected languages, and thus having only small variations (most notably in its declinations) from the root. This decision was made to mediate the differences induced by using several languages.

C.5 Operators

Operators (or symbols) were value-manipulation elements. They combined the value of the elements on both sides into a new value, that could be used for further manipulation or value comparison. Each of them corresponded to one of the following arithmetic operations:

1) Addition;
2) Subtraction;
3) Multiplication;
4) Division.

All operators were represented in a frame shown in Figure C.3, and in valid distributions shown in Figure C.4. Addition was represented by the plus "+" symbol, made up of two matchsticks, one laid horizontally and one perpendicular to it, crossing it in the middle. Subtraction was represented by the minus "–" symbol, made up of only a single matchstick, laid horizontally. Multiplication was represented by the multiplication sign "×". It could be seen as a plus sign tilted at 45 degrees. The division sign was represented by the forward-slash "/", which could be created with a single matchstick, much like the minus sign, but tilted at 45 degrees counter-clockwise.

FIGURE C.3: Operator frame

Matchsticks could be scattered in $2^4 = 16$ different possibilities within the operator-frame display, but only the 4 variations shown in Figure C.4 represented arithmetic operators and were considered valid.

FIGURE C.4: All valid operator representations
(top-left: addition, top-right: subtraction,
bottom-left: multiplication, bottom-right: division)

The operators inherently supported the order of operations where multiplication and division are evaluated before addition and subtraction. However, with a careful equation selection process, we made the order of operations effectively an obsolete rule (more on this in the section "No operator order priority (C.7.2)").

The visual representations of operators were selected based on their similarity with common arithmetic symbols and representational simplicity that can be achieved with virtual matchsticks. The multiplication operator is represented by the "×" symbol, commonly used in primary schools for multiplication arithmetic. The other common multiplication symbols in academia and finance sectors were considered, but not used. The dot "·" symbol was simply not representable with matchsticks, and the star "*" symbol turned out to have provided lesser clarity during the testing sessions.

In the instructions for the experiment we referred to these operators as "sign symbols". During the testing sessions, it turned out that the participants had a hard time understanding the term "operator" and its meaning in the context of an equation. Although this term would have been preferred for the same reason as "numeral", as described above, the participants reported that the term "sign symbol" provided better clarity of the designated concept, especially when used alongside the term "numeral", such as: "numerals and sign symbols". In the scope of the thesis, we continue to use the term "operator" for these value-manipulation elements, due to the term "symbol" technically denoting a much greater spectrum of concepts, which also includes the term "numeral" itself.

C.6 Comparator

The comparator element (or equal sign) validated the comparison between the left-side and the right-side calculated values of the equation. The equation was marked as solved when these values matched. Thus, it was an essential part of the equation and every equation had exactly one. In our experiment, it was represented by two horizontal matchsticks in the form of a common equal symbol as shown in Figure C.5.

Because it was a crucial part of the equation it made no sense for participants to remove or displace the used matchsticks. To avoid any issues that might have arisen in such cases, we had removed the option for participants to interact with the matchsticks used in the equal symbol. In this sense, the comparator element was always considered valid.

FIGURE C.5: Comparator frame and its matchstick-filled version

In the earlier versions of the experiment, we included the not-equals symbol, with a third matchstick crossing the equal symbol "$\neq$". This would negate the validity of an equation, marking it as solved when the left-side and right-side calculated values did not match. This element was removed because it made every equation too simple to solve. Similarly, symbols for non-equality ("<" and ">") were considered to provide an increased variety of options to solve the provided equations, but were later abandoned. They can be created by adding an inverse division symbol above ">" or below "<" another division symbol.

We also explored the option of combining the comparator with the operators, where removing one of the comparator matchsticks would create a subtraction operator. Such a combination increased the number of possibilities three-fold as it enabled every operator to become a comparator and vice-versa, thus completely destroying the equation frame (as described in the following sections). This provided too much flexibility for the participants as they would frequently create equations with zero or multiple comparators, getting confused in the process. Additionally, this model did not serve well with our model of designated mental-representation blocks. A single block would have to contain both operators and comparators, which turned out not to be the case, according to our test participants. For the sake of simplicity, this option was scrapped.

C.6.1 Element and equation frames

Matchsticks could not be positioned anywhere on the canvas. To create valid elements and equations, we have employed frames, i.e. distributed positions where the matchsticks can be placed. Each element type (numeral, operator and comparator) had a specific assigned frame, shown in Figure C.6. These frames served as placeholders, on which and only on which the matchsticks could be placed. Since the starting position of every matchstick was in one of the frames, we restricted matchstick move options exclusively from frame to frame (or within a frame), thus effectively quantifying the matchstick moves.

FIGURE C.6: All element frames
(from left: numeral frame, operator frame, comparator frame)

Equation frames are flexible structures with endless possible variants. For the purposes of our experiment, we were not interested in frame manipulations, so we had to select only one frame and stick with it throughout the whole study. The following restrictions were applied in order to make an informed selection of a frame used in the study.

Restriction 1

Numerals, operators and comparators have distinct frame elements.

Restriction 2

There has to be exactly one comparator frame per equation.

Restriction 3

Each operator and each comparator has to be surrounded by two numerals.

Restriction 4

No two numeral elements can be side-by-side.

Restriction 5

The comparator is in second-to-last place.

Restriction 6

Minimal requirements

Participants were able to move matchsticks within a frame, changing numerals into other numerals and operators into other operators, but were unable to transform a numeral into an operator or vice-versa in the same frame. This was a design decision, which was summed up in restriction 1 and was needed to facilitate the mental representational models described in this study. Without this restriction, the separation of frames based on element types (numeral, operator and comparator) ceased to make sense as these elements were able to transform into each other, providing undesired flexibility in manipulating and solving equations.

With restriction 1 in place, not all equation frames inherently facilitated an arithmetically solvable equation. For example, no valid equation could have two operators directly side-by-side. To achieve arithmetical solvability, we needed to add restrictions 2 and 3. This created an infinite set of possible solvable equation frames. There was one exception to this rule, a unary minus operator, which is a minus applied to negate its following numeral. We could have allowed the minus to be applied to the leftmost numerals on both sides of the comparator to simulate this operator. However, we decided that this was not really needed and could potentially confuse the less mathematically proficient participants; therefore, we opted out of using it.

Restrictions 4, 5 and 6 were specialized for our experiment in order to select a single equation frame and use only equations within that frame. That way we avoided possible confusion with different frames.

With restriction 4 we limited the experiment to single-digit values. It was simply not possible to input a value greater than 10. It was still possible to get a multiple-digit value through arithmetic calculations; however, this would also be somewhat restricted with the inclusion of restriction 5.

With restriction 5, by restricting the comparator to the second-to-last element place, we have ensured that all calculations were done on the left side of the equation. The test participants confirmed that we are commonly trained (in consideration of the school curriculum) to evaluate arithmetic operations on the left side of the equation and then write the result to (or compare the result with) the right side of the equal sign.

As specified mental representation block with inter-move matchsticks required at least two elements of each type. To facilitate these representation blocks, we required at least two numerals and at least two operators. With the assistance of restriction 6, we removed all frames with more than 2 operators and settled for a minimal viable solution.

After applying all of these restrictions, we were left with only one possible equation frame. In this frame, the element frames followed in order from left to right: *numeral, operator, numeral, operator, numeral, comparator and numeral*, as shown in Figure C.7. The sequence of these frames was fixed and did not change during the course of the study.

FIGURE C.7: The final selected equation frame

C.7 Equation selection

The matchstick equation selection process took several iterations, each of which restricted the pool of possible matchstick moves and possible solutions. Each iteration was carefully tested and checked for any possible errors.

The final set of equations was specifically selected for the experiment according to the following restrictions:

1) Each equation is solvable in one matchstick move;
2) Matchstick cannot be discarded; it can only be moved from one frame position in the equation to another;
3) Matchsticks cannot be moved one onto another;
4) Each move adheres to a strategy for up to two moves;
5) No operator order priority;
6) Only 2, 3, 5 are used;
7) Each initial equation is valid, but not solved.

Finally, the actual equations used in the experiment were chosen from the restricted equation-pool with an additional attempt to avoid obvious similarities between the selected equations, such as "2/2+2=2", "3/3+3=3" and "5/5+5=5".

The restrictions 4, 5 and 6-7 are a bit more complicated and will be discussed in detail in the sections below.

C.7.1 Each move adheres to a strategy

Restriction 4 was implemented in the tasks, so that the initial equation could not be solved by applying a matchstick move from a numeral to an operator or vice-versa. In other words, every move had to adhere to either strategy 'a' or strategy 'b' (a solution in two moves where strategy 'a' was applied once and strategy 'b' once, still adhered to this restriction). It was possible to apply this restriction for up to two matchstick moves. A stricter version with up to three moves or more, restricted the pool of available tasks too much and there were no tasks that would meet all the requirements.

C.7.2 No operator order priority

In the testing process, it turned out that the participants had problems remembering the order of operations, which requires that multiplication and division are evaluated before addition and subtraction. Their attempts at calculating the equation in the majority of cases consisted of an initial evaluation of the left operator and then an evaluation of the right one, which resulted in confusion when the participant tried to locate a flaw in their approach. In the actual experiment, such intervals of confusion could have led to lost time and thus corrupt the results. Unfortunately, some participants might have still employed the order of operations, so we were unable to simply remove this rule. However, with a careful selection process, we were able to select a set of equations where the equation's primary and secondary solutions evaluate in the same way, regardless if the participant was employing the order of operations rule or evaluating from left to right. In other words, if the second operator was a division or multiplication one, it would have never been preceded by an addition or subtraction one. This restriction only applied to solution equations; original equations were allowed to break this restriction since they were incorrect in the first place. Such a set of equations rendered this problem moot for the majority of participants. The participants who used three or more matchstick-moves in a particular task could still have experienced the problem, because there almost always existed a solution that broke this rule, but the participants doing so were few and we could safely disregard this problem.

C.7.3 Numeral restriction

There was a huge skew in favour of a numeral mental representation block. There were 10 possibilities for every numeral element and there were 4 numeral elements in an equation, while there were only 4 possibilities for every operator and there were only 2 operator elements in an equation. Operands occupied 1 to 2 matchsticks in 4 possible spaces, while numerals

2 to 7 occupied matchsticks in 7 possible spaces. While we could not state quantitatively how they differed, we could (qualitatively) see that in each of the three points above, the operands had a lower number of matchstick-position options compared to numerals. The sheer number of matchsticks used in numerals suggested an increased availability of the numeral representation (strategy ‘a’).

To mitigate this potential issue, we effectively restricted the numeral elements to only a subset, by selecting equations where the numerals used in the initial, unsolved equation and after the first step were in group {2, 3, 5}.

Discussion of this restriction

We have researched many different groups, but would argue that it is not possible to find a better group with only three elements, than the proposed group. In this group, there existed a relationship between elements that is hard to describe, let alone quantify:

- Two of the elements when added together made the third one;
- All of the elements had the exact same number of matchsticks;
- One element could be transformed into another with a simple move of a matchstick inside the element;
- No single move could be made to remove a matchstick from a numeral – a single-move solution with a move between numerals was not possible.

The other low-numeral groups that could support such a restriction were also {1, 7}, {3, 6, 9} and {1, 2, 3, 5}. Each of them had different problems and features, but we decided that the problems outweighed the benefits and therefore did not select any of those groups. We were also confident that this group would have the best number-of-possible-equations to group-size ratio.

Let us observe how this restriction affected strategy ‘a’. Remember: the strategies themselves did not change; we simply manipulated the tasks, so that the strategies could only be applied in a certain way! Using only one move, strategy 'a' had only 4 possibilities, as shown in Figure C.8.

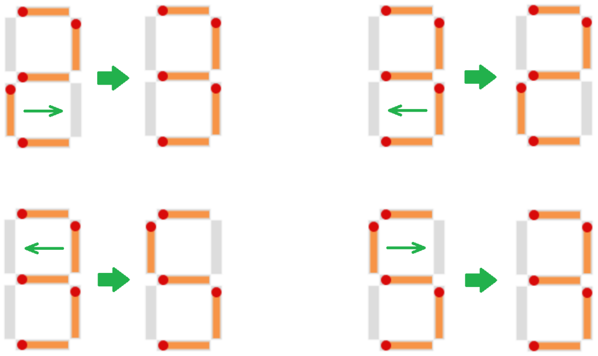


FIGURE C.8: All available single moves for strategy 'a' under numeral restriction

These were all only within-element moves: with one move it was not possible to move a matchstick from one numeral to another and create a valid equation. Using strategy 'a', it was not possible to create a different element than {2, 3, 5} in one move using any of the strategies. It was indeed possible in two moves (the possible elements were {6, 9} and rarely {4, 7}).

For example, one might be wondering why we could not create a 6 from a 5. Seems reasonable. But for that we would require an additional matchstick, that which we would have to introduce to this element. However, if we tried to take a matchstick from any other numeral (which were only {2, 3, 5}), we could not have made a valid digit from that numeral. Hence, other numerals could not be created in one-move only. Of course, a participant could still make all other possible moves, but in that case, they would have had to use more than one move to solve the equation.

Contrary to our efforts to curb the availability of strategy 'a', analysis of the results showed that strategy 'a' remained significantly more available than strategy 'b' even after numeral restriction had been applied.

Appendix D

Application design

D.1 Features of online experimentation

Executing an experiment online brings numerous benefits to a study in general. On the other hand, there are also a few drawbacks to this approach.

PROS

- More subjects
  It is much easier to acquire the needed amount of subjects.
- Minimized preparation time
  The experiment does not need to be set up / reset for every subject.
- No need for the experimenter to be present at the experiment
- Simultaneous experiments
  More experiments can run simultaneously.
- Experiments are more rigid
  Every experiment is executed in the same way in a predetermined and reviewed fashion. The experimenter cannot influence the course of a single experiment.

CONS

- There is no experimenter present to guide and verify the experiment.
  We have to trust the subjects a whole lot more.
- Much more preparation time spent on the application

After weighing the pros against the cons, we determined it was indeed beneficial to make an online application to perform our experiment. From here on, the software developed for the process of executing matchstick experiments is called "Matchstick application" or "the application".

The application was only made available for use on personal computers. No support was made for mobile devices, because of their different mode of input. Not only could touchscreen input have interfered with matchstick manipulation, but the participants would have to be split into different groups due to inconsistencies in their experiment participation. To avoid this issue, we blocked any attempt to participate in the experiment on a mobile device and kindly requested the participants to participate in the experiment on their personal computer. Additionally, this made it harder for participants to participate in the experiment multiple times by using different devices.

D.1.1 Application requirements

The application:

- Needed to be available on-line with a shareable link. The link was selected to be easy to remember.
- Needed to keep track of the subject.
- Needed to keep the subject's focus.
- Needed to be robust. All of the possible subject's actions needed to be accounted for.
- Needed to be able to properly cater to the experiment's needs.
- Needed to reliably store the gathered data

D.2 Database

In a locally run instance of PostgreSQL database, we stored the following information:

- Participant's sex;
- Participant's age;
- Participant's selected country *– might differ from the one obtained from the IP*;
- Participant's education;
- Participant's encrypted IP address *– used for tracking*;
- Participant's threat flag *– mark of possible malicious intent*;
- Assigned matchstick group;
- Individual data for each matchstick task:
  - Every individual matchstick-move in a task;
  - Usage of the restart button;
  - Cumulated number of moves used in a single task;
  - Total time used in a single task.
- Participant's comments.

This data was not sufficient to fully determine the identity of a participant. Nevertheless, it was secured and only made available to the administrators of the experiment.

D.3 Domain name

The domain name is the initial URL address one enters in the browser bar in order to access the desired webpage. The domain name should be simple, consisting of one or two words combined in the English language, which should encompass a special important segment or a general idea about the experiment. Such a domain name enables subjects to easily remember the application site and enables verbal sharing and subject acquisition.

We decided on the domain name `matchstick-task.eu`, as some other ones, such as *matchstick.eu*, were either too expensive or unavailable.

D.4 Server

The application and the database were deployed on a publicly accessible server with a static IP. A simple publicly hosted server of the Czech company Hukot was rented for this purpose and registered with the public IP 46.36.38.196.

D.5 Keeping track of a participant

The application was running partially on the subject's device to avoid performance issues and any possible lag. However, this created more potential data collection issues as we had to put our trust in the participant's device. For example, if the participant lost their internet collection, if their browser crashed, or if the computer powered down, we provided support for their device to properly recover and continue with the experiment. Experimental data was sent in regular intervals after every task, to minimize potential loss of data. All of the traffic between the participant and the server was encrypted for the participants' protection.

To achieve high anonymity, we have abolished participant verification. At no point in the experiment was the participant required to provide their crucial personal data, which would enable us to verify and identify them as an actual person. However, we still had to enforce some form of partial identity, so that we were able to assign different activities to a single participant. In this effort, we introduced a mechanism to track the participant's device, which was much more affordable, since it was already implemented for network activities. Each device was, for the purposes of this experiment, identified with one specific person. The term "participant" in this thesis is loosely related to a single person-device pair. Cases where this was not true, where multiple people used the same device or a person used multiple devices, are described in the next paragraph.

A series of checks was implemented to track a participant every step of the way. The frontrunner in this effort was a browser cookie. Upon first connection to the website, the participant is sent a "cookie". A "cookie" is a special identification value that is sent with every future network request – all of the network requests pertaining to a single participant are accompanied by this value. Similarly, all of the received data was stored in relation to this value, enabling us to identify all actions of this single participant. The cookies resided on the participant's device and could have, unfortunately, been easily removed by them, either intentionally or unintentionally. For this purpose, we also tracked the participant's IP address on the server side. All network requests received from a single IP address were associated with the same participant. IP addresses were never stored plainly in the database; instead, we stored encrypted versions of those IP addresses and compared them amongst themselves, providing an additional layer of anonymity. If one of these mechanisms failed or produced inconsistencies, the participant was flagged as potentially malicious and their experimental data was subjected to a manual review to identify potential malicious content. Similarly, if there was any suspicion that the data was not generated through the actual application, but instead written or edited elsewhere, the participant was flagged. Such data included unreasonably short reported task duration, inconsistency of matchstick moves, or repeated attempts at single tasks.

Unfortunately, it was impossible to cover and avoid all possible malicious attacks, while trying to maintain such a level of participant anonymity. If a person used various personal computers, or cleared cookies and changed their IP address (for example with VPN services), we were unable to recognize them each time as the same participant, thus allowing them to attempt the experiment multiple times. To curb such attempts, we implemented an easier and more lucrative way for malicious participants to retry the experiment. The experiment could have easily been retried on the same device multiple times. However, all subsequent attempts at it were flagged and their experimental data excluded from the final data analysis. Through the repeated experiment, participants were unaware and unable to determine that their session had been flagged and thus would not be considered in the analysis.

Online security is a best-effort science. There is no perfect protection that would eliminate all potential attacks 100% of the time. Regardless, computer scientists make it harder and harder for attacks to be executed, which is sufficient for normal everyday usage. This design was created in the hope that attackers would take the easier way and attempt their attacks through a path that was, unbeknown to them, monitoring for such activity. This maliciousness-control mechanism unfortunately also flags multiple individuals who attempted the experiment on the same device. For the sake of security, we had to disregard their data as well. During the data-collecting phase, we received information on specific people, most notably close friends, where the entire family participated on a single device. In such cases, we made notes of the time and IP addresses of those participants and, if we were able to determine which attempts were theirs, allow all of those multiple attempts to be used in the analysis.

D.6 Task page design

D.6.1 Moving matchsticks around

When the matchstick was on the move (removed from its initial location, but not yet placed in a new one), some form of indication was needed. For these instances, we implemented a matchstick image shown on the tip of the mouse, so the participant actually feels like they are moving the matchstick.

Matchsticks could only be placed within designated frames to avoid invalid states (e.g. a participant tried to create a numeral 4 with a diagonal matchstick). This way we could suppress thoughts about invalid moves and make participants focus only on the actual valid ones.

The participant's activity was also kept to a minimum in order to eliminate the effect of those actions on the task. For example, if the participant had to click to turn the matchstick around, the number of clicks required could potentially influence the user's decision. To prevent this, we had introduced an automatic close-frame association mechanism, which popped a dragged matchstick into the closest available position within a frame when dragged sufficiently close; in this way, it showed the participant what their current activity might result in. This way, the participants did not need to burden themselves with the problem of properly positioning the matchstick that they wanted to place.

A participant could have changed their mind about the matchstick taken - if the said matchstick was dropped back to its initial position. This move was logged, but disregarded in the analysis phase.

Regarding the approach to the move-action, we have identified two groups of participants, each with their own style of matchstick manipulation:

1. Representation-grab: this group contains participants who think of matchstick-moves in terms of removing and then adding the matchsticks / picking them up and placing them back down.
   A real-world example is picking a matchstick up vertically from the table, holding it in the hand, and placing it down on the table.
2. Representation-drag: this group contains participants who actually view a move as a single unit, dragging a matchstick along the surface of the canvas.
   A real-world example is dragging the matchstick from its initial to its

target location along the surface of the table, without the matchstick ever moving in the vertical direction.

When implementing the experiment to resemble human behaviour in the experimental set-up, we tried to encompass each of the mechanisms of matchstick manipulation that were available to participants in the physical, real-world version of the task (e.g. a task with actual matchsticks that needed to be moved by hand). Upon careful examination of different possible move implementations, which are generally used in similar simple computer games, we observed two different mechanisms for performing a single move. We felt that each of them related to a different above-mentioned style, however, further investigation is required. These action implementations were:

1. Action-grab: the participant clicks on a matchstick to take it from the start position and then clicks again to put it in the end position. This mechanism corresponds to representation-grab.
2. Action-drag: the participant clicks and holds to move the matchstick from the initial to the target position, where they release the mouse button. This mechanism corresponds to representation-drag.

Both mechanisms were enabled in the experiment and participants could use whichever they preferred. They were also both considered highly intuitive, so no effort was made to present them and their differences to the participants.

D.6.2 Colours

Colours were selected to be as simple as possible, so as not to disturb the participants. We decided to go with a standard light-brown (#F79447) matchstick stick and a red (#D80A0A) matchstick head, as such matchsticks are most common in local shops (Slovenia, Slovakia and Austria). Buttons were made a in light shade of blue (#378de5), as we felt it suited the matchstick task design. A special exception was the failure-success-button, which used bright colours to clearly express the failure and success state with red (#fe1a00) and green (#5cbf2a) colours, respectively. On the blue background (#378de5), the text was mostly written in a white (#FFFFFF) Arial font, for readability purposes. Likewise, on white background surfaces, the text was written in the previously mentioned blue colour. Aside from the matchsticks and the failure-success-button, we used only the aforementioned light blue and white colours on the task page to preserve coherence. A good idea for a future study might be to examine potential colour effects on the participants' performance in the experiment.

D.6.3 Translations

We provided support for 4 languages:

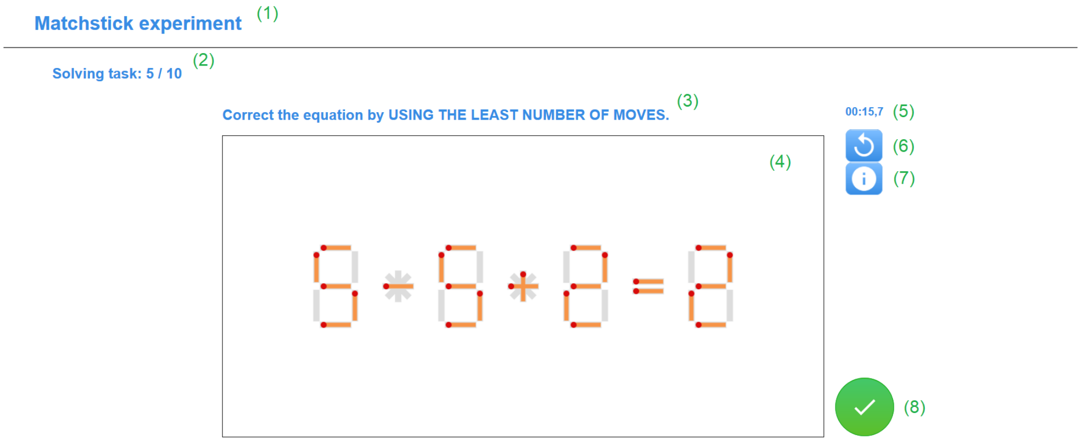
- English;
- Slovenian;
- Slovak;
- German,

The language could be selected on the very first webpage with very intuitive country-flag buttons, each representing its related language. The language could also be changed at almost every step during the experiment with a small country-flag button on the top-right side of the page. Every piece of text displayed on every webpage had been manually translated to all 4 languages. Each translation was been checked by professional translators or native speakers of that language.

D.6.4 Task page layout

The task page was a simple webpage, which became accessible after the user provided some personal data and accepted the experiment terms and conditions. The page differed between experiment phases; in phase L2, the equation canvas was replaced by a video and in phase L1, the whole page was overwritten to provide information in a readable format with visual support.

Here, we discuss in detail the structure and elements on the page in the testing phases. All elements are referenced in Figure D.1 by the attached numbers in parentheses.



Matchstick experiment

Solving task: 5 / 10

Correct the equation by USING THE LEAST NUMBER OF MOVES.

00:15,7

FIGURE D.1: Example of a task page
including numbers in parentheses for in-text explanations

Experiment title (1)

Simple experiment title.

Task counter (2)

Counter, displaying the current task number and the total number of tasks. It helped participants to track their progress through the experiment. Before this counter was added, the test participants had reported reluctance to follow through with the experiment, as it had not clearly shown the progress and end of the experiment in a series of fairly similar tasks.

Task instruction (3)

An additional phase-specific instruction and directive that informed the participant of the task goal in the current phase. It read *“Correct the equation by USING ONLY ONE MOVE”* in phases T4 and T8, and *“Correct the equation by USING THE LEAST NUMBER OF MOVES”* in phases T5, T6 and T7.

Equation canvas (4)

This was the area where the matchsticks were allowed to be moved and interacted with. All matchstick- and equation-related activity was restricted to this area. Matchsticks could be picked up and dropped with a single mouse click and dragged across the canvas.

Timer (5)

A timer displaying the accumulated time spent on the current task. The time display was up to 100 milliseconds accurate and served as a constant reminder for the participant to solve the task as quickly as possible. The timer was reset for every task.

Restart button (6)

The restart button enabled a participant to reset the task to its initial unsolved state. In the test runs, it turned out that some participants were more visual thinkers and liked to move matchsticks around while they explored possible solutions. In some cases, they got lost and could not re-trace their moves in the opposite direction. The restart button proved to be a useful feature in helping them to get back on track. Additionally, we were not interested in the solutions which took many moves to achieve, but rather in the simple few-move solutions where the strategy usage was clear. This button only reset the task state and did not affect the time or move counters.

General instructions button (7)

This button enabled a participant to revisit the general task instructions page, which was displayed in phase L1. It opened the fallback information page in case a participant had skipped or forgotten the information provided in phase L1. When the information was displayed, the timer did not stop, which was also clearly stated in the said information.

Pause button (removed before conducting the experiment)

In the initial versions of the experiment, we provided a pause button, which would have allowed participants to temporary pause the execution of the experiment. The motivation for this button was that in case of emergencies that, for some reason, required their immediate attention, participants would likely stop their participation in the experiment. In case they later returned to continue the experiment, the time data would have been corrupted (including the time spent on solving the equation and the time spent on dealing with the emergency) for this particular task and we could salvage the data in such cases. However, the button could have easily been abused for participants to take breaks and disrupt the flow of the experiment. Not only would the participants have time frames during which they could forget the learned strategies, but the participants might not have the same amount and duration of breaks. In the end, we decided to remove this option, because it could severely harm the experiment; the cases where participants took emergency breaks, were simply filtered out.

Fail/Success indicator & Continue button (8)

This button served as an indicator of whether the equation was solved. The equation was validated after each match move and determined whether it was solved or not. When the equation was not solved, the button was red, with a white cross in the middle, and nothing happened when it was clicked. The moment the equation was solved, the button turned green with a right-pointing white arrow. When the green button was clicked, the participant proceeded to the next task in the series.

D.7 Online support

The application has been adapted to work in the following browser versions:

- Mozilla Firefox, 74.0 (64-bit)
- Chrome, 80.0 (64-bit)
- Safari, 5.1.7
- Opera, 67
- Internet Explorer, 11.718

As mentioned before, mobile devices were prohibited from the experiment altogether.

Appendix E

Possible improvements

During the data-acquisition period, we received some comments and suggestions from the participants regarding our experiment. Participants were able to write their comments anonymously in the forms provided right after finishing each experiment. There were some participants who were unable to finish and they shared their point of view on the matter. On the other hand, we received some comments directly by talking to the participants who had confirmed participation in the study. We discuss the most important ones below and provide our opinions and comments about them. All in the hope that they can be of help in similar future experiments.

E.1 Bad education selection

A few participants complained that there were not enough education options available, which would suit their educational background, especially the "engineers". This is a known problem and was already discussed in detail in Appendix B - Sensitive information.

E.2 Participants got stuck

Most of these reports came from a verbal discussion, as almost all the participants, who got severely stuck, were unable to finish the experiment and gain access to the comment field. They complained that no help was provided and since they were unable to continue with the experiment, they simply quit altogether.

This is a known issue and was deliberately left unhandled. We simply could not provide any sort of hints or additional information to the participants, as that would invalidate the data collected from those sessions. The decision was made to simply give participants unlimited time and let them explore various possibilities of the matchstick-task. The reasoning behind it was that, if the participant was given no help or no option to skip the task, and on the other hand, given no time-limit, they might focus exclusively on the task and perhaps on completing it, even if it took longer. There are countless possible solutions for each task, so they should have been able to reach one eventually. On the chance that they were unable to do so, they could just quit the experiment, as the collected data could not be used anyway if hints were involved. This way, in the worst-case scenario, we simply got no data from the participant. However, as it turns out, this brought even worse results than we had expected: it created a bad public opinion of the experiment. The participants who got stuck were unhappy with the experiment and were unlikely to recommend it to others. Furthermore, some actively discouraged others from participating. Luckily, the damage had not been severe.

With this newly obtained knowledge, we argue that an option to prematurely quit the task or experiment would be a good benefit, but this option should not be immediately available. Our suggestion is that after a predefined period of time (a good estimate is 5 minutes), the participants could be given a button to skip the problematic task or to end the experiment altogether. Of course, the participants have to be informed of this option. With this option in place, we argue that a lot of negative publicity could have been avoided.

E.3 Not understanding the available symbols

Participants were never specifically informed of all the available numerals and operators. In all the preliminary tests, it seemed that the symbols used were clear enough on their own and that the test-participants did not need any help with them. However, it turned out that some participants did have problems, because they did not know what symbols were available. Especially problematic were the division operator “/” and the numeral four “4”.

With this problem in mind, we propose two potential improvements:

- Participants should be presented with all of the possible elements used in the task beforehand. This list should contain all the numerals and all the operators in both the visual and the arithmetic representation.
- Each participant should get a short training session consisting of two steps. In the first step, the participant should train in matching the visual representations of organized matchsticks with their corresponding arithmetic value. In the second step, they should evaluate short partial equations in an arithmetic manner (e.g. "6-3" and they have to select the value 3).

These suggestions have been taken from Knoblich et al. (1999).

E.4 Clunky matchsticks

A few complaints addressed the matchstick manipulation. Some participants found it difficult to position the matchsticks in their desired location within the operator frame. This problem had not been observed in the testing process. In order to reduce the participants' frustration, it might be useful to add a short training session before the experiment, teaching the participants how to manipulate matchsticks within operator frames. However, this is not a severe issue, as all the participants who complained, have successfully finished the experiment.

E.5 Different screen sizes

Some participants complained about matchsticks being too big or too small. However, they are somewhat dependent on their screen size and resolution. The experiment page can be zoomed in or out, like all other webpages in any other browser, but it seems not all participants were aware of that browser option. However, adding additional zoom buttons would be an overkill, in our opinion.

E.6 Hard-to-see elements

Some participants complained about the matchsticks being too vivid and that it was hard to see the arithmetic numbers. This is actually part of the chunk-decomposition design. The participants should have to switch between representations: (1) to visual representation in order to change the equation and (2) to arithmetic representation to validate the equation.

Some others complained about the colours of the matchsticks making them feel dizzy.

There is nothing to be done regarding these two problems.

E.7 Participants' scores

A lot of participants expressed an interest in their own scores and their relation to the scores of others. We suspect that they did not fully understand the concept of the investigative experiment and instead thought of our experiment as some form of personality or IQ test (as noted in their remarks, e.g. "So, when do I get the results of my IQ level?"). Even though that is not required, we should have shared some information with the participants. We assume that their task times and moves, compared to average times and average moves, would be sufficient information. We assume that this would have satisfied the participants and reduced the number of requests we have received asking about their scores.

E.8 Getting in contact with participants

We suggest adding a field for participants to provide their e-mail address. A lot of participants in the comment field requested "their personal scores" (discussed above) or additional information about the study and its results when they become available. However, most of them did not include any way of contacting them. Since the experiment is completely anonymous, we simply cannot track those participants down and fulfil their requests. An additional warning should be added to the new e-mail field, alerting participants that providing an e-mail address breaks anonymity.

E.9 Hacks

Some participants reported possible bugs and hacks in the system. Especially regarding abusing the application and participating in the experiment multiple times. They reported these "hacks" without knowing that their subsequent attempts get flagged and do not actually count towards the experimental data. However, these participants did not abuse their newly found “hacks” and it is always a good thing to get feedback on such potential issues.

In conclusion, we have observed a variety of participants. Some complained that the experiment was “too easy”, while others marked it as “incredibly hard”. There were also some participants who figured out that there was some form of priming involved, informing them that they can only move certain matchsticks. On the whole, we received a very mixed set of participants with a variety of experiences with our experiment. Their overall response is decisively positive, as the participants were mostly intrigued by the novelty of this experiment.