“HEY ROBOT, WHAT DO YOU THINK?”

HOW CHILDREN LEARN WITH A SOCIAL ROBOT

Daniel Patrick Davison
This dissertation has been approved by:

Supervisor
prof. dr. V. Evers

Co-supervisor
dr. ir. D. Reidsma

The research reported in this dissertation was carried out at the Human Media Interaction group of the University of Twente.

The research reported in this dissertation was supported by the EU FP7 Programme (FP7-ICT-2013-10) as part of the EASEL project under grant number 611971.

DSI Ph.D. Thesis Series No. 21-003, ISSN: 2589-7721
Digital Society Institute, University of Twente.

SIKS Dissertation Series No. 2021-06
The research reported in this dissertation has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design: Judith van Stegeren
Printed by: IPSKAMP, Enschede
DOI: 10.3990/1.9789036551588

© 2021 Daniel Patrick Davison, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.
“HEY ROBOT, WHAT DO YOU THINK?”

HOW CHILDREN LEARN WITH A SOCIAL ROBOT

DISSERTATION

to obtain
the degree of doctor at the Universiteit Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board
to be publicly defended
on Friday 30 April 2021 at 14.45 hours

by

Daniel Patrick Davison

born on the 10th of October, 1987
in Assen, The Netherlands
Graduation Committee:

Chair / secretary: prof. dr. J.N. Kok
Supervisor: prof. dr. V. Evers
Co-supervisor: dr. ir. D. Reidsma
Committee Members: prof. dr. D.K.J. Heylen
                      prof. dr. W.R. van Joolingen
                      prof. dr. ir. P. Markopoulos
                      dr. F. Tanaka
                      prof. dr. F. Eyssel
                      prof. dr. ir. H.J. Hermens

Paranymfs:

Arne Bakker
dr. Lorenzo Gatti
Summary

This thesis explores how a social robot can have a meaningful impact on the learning process of primary school children. As robots become more available and accessible, we expect that they will gradually move into the classrooms of the future. Before such a transition can take place we need to better understand in what ways robots may excel and add value to the learning process. We also need to gain a better understanding of what it takes to develop and use them in real classrooms for longer periods of time.

We investigate learning situations where a single child works on a learning task in the presence of a robot. Our tasks are based on inquiry learning principles, in which children explore a certain phenomenon by conducting small controlled experiments. Throughout this structured process of exploration they can benefit from additional external support from interactive learning technologies. The work presented in this thesis shows how children's learning process can be positively affected by delivering certain forms of social support through a robot.

We argue that a robot—by virtue of being a robot—is in a unique position to offer these kinds of support. We consider the robot as a holistic entity consisting of many social features and modalities, together establishing it as a complete and convincing social agent. To show the effects of the robot as a whole, we compare it to situations where children learn without a robot; support is then delivered through a less-social tablet device.

To inform the design of our robot's social behaviours we observed pairs of children learning together. Concrete design guidelines were derived from the observations using a grounded theory approach. Using the guidelines, we developed and evaluated four variants of our Robot-Extended Computer Assisted Learning (RECAL) system. When working with our robot, children gave better explanations and showed an improved mindset towards learning—robots can make a meaningful contribution!

Furthermore, we share insights to gain a better understanding how technologies like our RECAL system may be used for conducting research in the wild over longer durations. We show how our autonomous robot and learning tasks offered sufficiently rich and challenging content to elicit unsupervised recurring spontaneous interactions with children throughout a four month period. We also discuss how in-task experience sampling and embedded sensors in the learning materials were used to follow their progression through tasks and difficulty levels. With these first steps we have shown that conducting long term studies in the wild is a feasible endeavour yielding valuable insights. To advance the field of HRI and move our robots out of the lab into the hands of real users in real contexts, it is absolutely necessary that more of such research takes place.
Deze thesis vekent hoe een sociale robot een betekenisvolle bijdrage kan leveren aan het leerproces van basisschool kinderen. We verwachten dat naarmate robots gaat-deweg goedkoper en toegankelijker worden, zij steeds vaker gebruikt zullen worden in onderwijs. Voordat een dergelijke transitie plaats kan vinden moeten we beter begrijpen waar robots in uitblinken. We moeten ook een beter beeld krijgen van de uitdagingen die te wachten staan wanneer we een robot ontwikkelen die langere tijd in een echte klas moet kunnen werken.

We bekijken leersituaties waar een enkel kind in het bijzijn van een robot aan een leertaak werkt. De taken zijn gebaseerd op onderzoekend leren, waarbij kinderen een bepaald fenomeen verkennen door het uitvoeren van kleine experimentjes. Interactieve leertechnologie kan kinderen allerlei ondersteuning bieden gedurende dit proces. Het werk gepresenteerd in deze thesis laat zien hoe een mate van sociale ondersteuning door een robot een positieve bijdrage kan leveren aan het leerproces.

Wij stellen dat een robot—omdat het een robot is—een unieke rol heeft om dergelijke ondersteuning te bieden. Bij beschouwen de robot als een holistisch geheel met een veelvoud aan sociale eigenschappen en modaliteiten waarmee het communiceert. Om de effecten van de robot als geheel te onderzoeken vergelijken we het met situaties zonder robot; ondersteuning wordt dan geleverd door een minder sociale tablet.

Als inspiratie is goed gekeken naar hoe kinderen elkaar ondersteunen. Op basis van observaties zijn concrete ontwerprichtlijnen opgesteld voor de robot, die we hebben uitgewerkt in vier variënten van ons educatieve robot systeem. Kinderen die met de robot werkten konden hun redenatieproces beter uitleggen en lieten daarnaast een groei zien in hun mindset jegens leren—robots kunnen hierin dus een impact hebben!

We bespreken ook inzichten over hoe dergelijke technologie gebruikt kan worden voor onderzoek over langere periodes in echte klaslokalen. We laten zien hoe onze autonome robot en leertaken voldoende rijkheid en uitdaging bood om gedurende vier maanden terugkerende spontane interacties met kinderen te onderhouden, en hoe we dit proces subjectief en objectief konden volgen door middel van geïntegreerde vragenlijstjes en sensors in de leertaak. Hiermee hebben we de eerste stap gezet om aan te tonen dat het doen van dergelijk onderzoek niet alleen haalbaar is, maar ons ook waardevolle inzichten kan geven. Om verdere vooruitgang te boeken in mens-robot interactie zullen we vaker met onze robots uit het lab moeten treden om deze in de handen te leggen van echte mensen in de echte wereld.
Acknowledgements

Now is the moment where I finally get to look back at my (rather long, oops..) period as a PhD student and the lovely people I shared this journey with. To everyone who has supported and helped me, who I've loved, laughed and cried with, or have gamed, tinkered, gotten drunk and partied with, who have put up with my nonsense, corny jokes, and sarcasm, who I've bitched, moaned, and vented frustration to... and anyone who has otherwise been important in my life over the past years: Thank you all!

I’d like to start by expressing my gratitude to Dennis and Vanessa. You have been a great source of inspiration! It was always comforting to know I could count on you and I am grateful for your understanding and support through the ups and downs over the past few years. I feel I have gotten to know both of you on a personal level; thank you for your human-centred approach to supervision.

This is a terrible cliché but it rings true: Dennis, I could not have completed this journey without your guidance. Thank you! More than 10 years ago you supervised my bachelor project, sparking my interest in socially interactive technology and cementing my choice for following the HMI master programme. Later, while I was finishing up my Master’s thesis, I remember you poking your head round the door to mention a “cool project with robots”. Can’t argue with that, thus the ball started rolling and I stuck around for a while longer! I could not have wished for a better person to supervise me in becoming a doctor, thank you for always being available and for sharing many lessons in work and life. Speaking of which, I am constantly amazed and impressed with your ability to enjoy a rich family life with your wonderful girls and wife, and somehow combine it with an extremely busy work routine. Even as I am writing this, in the middle of the night, you are still giving helpful feedback and suggestions... You know at some point we must talk about a healthy work/life balance, but I will tactically wait until this book has made it to the printer! Throughout my PhD I have learned so much from you about research, computer science, and personal development. You have a way of swooping in at the right time, making a few on point (often sarcastic) remarks, and casually solving a problem I have been stuck on for days. I fondly think back to the many demos, school trips, and project integration meetings we organised and participated in together. Furthermore, our deep discussions where we dissected a particular technology, explored new theoretical models, and laid foundations for papers have been invaluable. Through this all you have guided me to become an independent researcher. Now, I often find myself (subconsciously) borrowing teaching and supervision techniques that I picked up from you. With varying degrees of success, I might add; I obviously still have a lot to learn and hope to continue doing so together.
Vanessa, as my promotor you have gently steered me in the right direction, ensuring my work was of good quality and relevant within the broader HRI community. At times this meant really zooming in to get minute study details right, or looking through collected data sets together to uncover and discuss interesting results. Other times it meant taking a few steps back and allowing me to figure things out, all the while keeping a keen bird’s eye view of my work. You also helped put my work in the spotlight, even if I was hesitant to do so myself. This sometimes meant introducing me to the rock stars in our field, and volunteering me for demos, talks, and interviews on tv and radio. I look back fondly on these experiences, even if they scared the crap out of me at the time! I’ve always enjoyed discussing my ideas with you, and appreciate the time and effort you took to offer feedback and suggestions. Thank you for the support and freedom to pursue my own interests and research directions throughout my PhD!

I am grateful to my paranymphs Lorenzo and Arne, you have been good friends to me over the years and I am honoured to have you (virtually) by my side.

I consider myself very lucky to have been a part of the HMI group for so many years (and still counting). You are a wonderful group of awesome people that inspire happiness and creativity! The many social activities like (board)game nights, MTG drafts, sports tournaments, DnD campaigns, BBQs, drinks, and group outings make it hard to imagine that actual work takes place. Thankfully, there are also much, much more serious meetings like the Thunder Thursdays, where we expand the boundaries of science and technology by doing stuff with things.

Betsy, a long time ago you helped me take my first steps as a researcher. Now, even after your retirement, you continue to help improve my papers and thesis chapters with your keen eye for methodology and statistics. I’ve come to know you as a warm and nurturing person and it is always a pleasure catching up about life inside and outside of HMI. Lynn, thank you for bringing a British touch to the department! You’ve been a lifesaver when proofreading my papers that needed a pinch more native English, thank you for all the helpful comments and feedback. I’ve enjoyed our coffee breaks to share the latest project gossip; your dry British humour combined with the occasional Dutch directness make for a great colourful combination. Charlotte en Alice, it’s always nice to catch up with some small talk and share stories about holidays, photography, baking, gardening, chickens. Thanks for being the backbone of HMI and helping everyone find their way around the bureaucracies of the university (and thanks for sneaking an extra meeting or two into Vanessa’s busy agenda from time to time!). Jelte, you are one of the most persistent and persuasive people I know; who would have thought that inviting you to play a few rounds of squash one day would lead to playing with half our colleagues??! Kidding aside, you’re a good friend and great listener and I sincerely thank you for our many personal talks about life, and for offering a good ear when I needed to rant about stuff.

Daphne, Jered, Alejandro C., Dees, and Judith W., thanks for all the laughs and discussions through, between, over, and around our monitors, you’ve been awesome office-mates! People who have been in and out of the ‘social robots club’ over the years, Manja, Daphne, Dennis, Kriet, Bob, Vicky, Maartje, Roelof, Cristina, Marty, Eduard, Gwenn, Jaebok, Michiel, Jorge, Jamy, thank you for all the nice meetings, keeping up with current events, discussing random ideas, practising presentations, and sharing feedback. Fellow winners of the HRI2017 Student Design Competition, Jered, Cristina, Jan, Jeroen, I wish you many more nights of refreshing, non-euthanised sleep! Bob, Jered, Jeroen, Jelte, Merel, Michel, thank you for introducing me to DnD and granting
Acknowledgements

+2 nerdiness. Cristina, your strong Italian perspectives and outbursts have been a great source of insight and entertainment. Jan, thanks for your humour and for the awesome side-project collaborations. Bridgette, thank you for our nice conversations, and thank you for sharing your mum’s traditional Bobotie recipe, we still regularly make it (and then eat for 4 days)! Dirk, your (often unexpected) philosophical remarks and propositions have been a welcome contribution to the random lunchtime discussions. Cristina and Roelof, Lorenzo and Iske thanks for the gezelligheid, can’t wait until we can meet up again for games, drinks, and dinners! Bob, of all the possible days in the year we happen to pick the same day to defend our thesis... Are we so predictable after all??

Finally, special thanks to Mariët, Gerwin, Sara, Ella, Simon, Anne, Jan v.E., Shenghui, Thomas, Angelika, Gijs, Juliet, Randy, Robby, Judith v.S. Deniece, Camille, Natty, Alejandro M., Merijn, Dong, Mannes, Dasha, Carolien, Birna, for the nice conversations and collaborations, and joining in games, laughs, coffee/tea breaks, and all the other things that make HMI amazing!

Many thanks go to the colleagues in the EASEL project, I’ve had a great time working with you all! Vicky, Maria, Pedro, Riccardo, Sam, Emily, David, Lorenzo, Daniele, I fondly look back on the many great BCBT summer schools, integration weeks, and technical demo preparations.

Within the UT-EASEL team we were lucky to have a rewarding multidisciplinary collaboration. Vicky (no no.. the other Vicky, you know, the Greek Vicky ;-) thank you for the nice discussions about how our work aligned with the theories on child development, this helped put things in perspective. Please also send my best wishes to your kids, over the years they were often the first to offer valuable feedback on our prototypes! Jan, thank you for keeping an open mind during the many in-depth meetings and brainstorms as we explored yet another crazy idea. Through our discussions your expertise on agents and interactive learning environments has helped shape the pedagogical fundamentals of our studies. Frances, thank you for all the fun times! Whether it was designing and building the learning tasks and the robot behaviours, setting up and running the experiments, sitting in schools pressing buttons and interviewing a seemingly endless line of participants, analysing results, writing the papers and the many revisions and rebuttals... The studies that are presented in this book were really a joint effort, it has been a pleasure collaborating with you.

To the colleagues I collaborated with in COUCH, Randy, Gerwin, Jan, its been a lot of fun to work with you and so rewarding to eventually see the Agents United platform come into existence. Furthermore, thank you to DE-ENIGMA colleagues, Pauline, Bob, Jamy, Bridgette, no matter what, we made an awesome team!

There are many students who I have happily supervised and worked with over the years. I have learned a lot from you, and I continue to be impressed with your creativity and effort and the crazy/cool outcomes of your projects. In particular I want to thank Louisa and Andrea who’s work is featured in Chapter 5, and Emiel for helping out with hardware and electronics at various stages of prototyping.

A huge amount of thanks goes to the many schools and daycare centres that opened their doors to us, and to the teachers who welcomed us into their classrooms, often disrupting the schedule, causing chaos, and distracting the children. Yet I am most grateful for the children themselves, who played with our robot, explored the learning tasks, expressed their honest thoughts, doubts, concerns, and suggestions, thus helping us better understand what this all means to them. This work would not be possible without you!
Acknowledgements

Lieve Badeendjes der Calslaan 22, bedankt voor alle gezelligheid en lol over de jaren heen! Ik denk met plezier terug aan de vele bizarre huisavonden, stapavonden, late-night game/movie-marathons, badjas-uitbrak-dagen, het vele kattenkwaad, de zellweekenden, kerstdiners, huisuitjes, ... Tip van een veel te oud fossiel voor de jonge feutjes en helden: neem je tijd en geniet er nog maar uitgebreid van, voor je het weet zit je tot je nek in de poepluiers ;-)  

Niels, my partner in crime, we gaan ver terug zeg, damn! Vroeger, toen het gras nog groen was, toen internet nog via postduif ging, toen dubbelvla nog gewoon in twee pakken kwam en je poep nog met een lange oe schreef, dat waren de dagen! Van stampen op Qlimax tot gillend over straat rennen, we hebben wat moois beleefd in die goeie ouwe tijd. Bedankt daarvoor! Ook meer recent tijdens en na je afstuderen, bedankt voor je onverwachte bezoekjes en de vele koppen koffie om de werkdag te doorbreken en gezellig even bij te praten.

Arne, wat zijn we toch altijd fanatiek met z’n tweeën; squashen, mountainbiken, wintersporten en na afloop een welverdiend koud bier, mij hoor je niet klagen! Ook op digitaal vlak hebben we het nodige gemeen en het is altijd weer mooi om lekker over technologie te ouwehoeren. Zonder gekkigheid, ik kan altijd alles aan je kwijt—Arne, bedankt dat je zo’n goede vriend bent! En natuurlijk kan ik Mirel daarin ook niet vergeten, hartelijk bedankt voor de vele gezellige avonden, lekkere dinertjes, drankjes en bordspellen samen.

Bas en Hester, bedankt voor de gezellige weekendjes en dagjes weg. Van uitwaaien aan het strand gaan we straks naar uitwaaien in de haverlucht, we kijken er naar uit!

Alle ooms, tantes, neven en nichten in het hoge noorden, bedankt voor de vele gezellige BBQs, feestjes en familie uitjes, en meer recent natuurlijk de diverse edities van de familiequiz!

Opa en Oma, mijn grootste fans! Over de jaren heen werd elke kans aangegrepen om aan Jan en alleman te verkondigen hoe trots jullie waren op de kleinzoon aan de universiteit. Altijd als ik op bezoek was waren jullie geïnteresseerd en benieuwd, "Hoe gaat het met je robots? Ben je al bijsna afgestudeerd?" Ik praatte er altijd met liefde over met jullie. Nu het eindelijk zo ver is, doet het des te meer pijn dat ik dit moment niet meer met jullie kan delen! Ik heb altijd genoten om bij jullie op de boerderij te zijn, bedankt voor alle mooie herinneringen!

Ook de familie in het wilde westen wil ik hartelijk bedanken voor de gezelligheid, gastvrijheid en steun. Hoewel middenin de grote drukke randstad kan ik bij jullie altijd ontspannen en tot rust komen. Bedankt Rob, Annemieke, Maaike, Suzanne, Oma Bep, Opa Klaas en Oma Toos!

To our family across the channel, it has been lovely seeing you all over the years! Infrequent and brief as our visits may be, you’ve always opened up your homes and hearts and taken excellent care of us. Nana, we miss you! Hopefully we can see you again in person soon, keep your spirits up!

Mum and Dad, thank you for the great childhood! While growing up we frequently travelled and experienced all kinds of exotic places and cultures, which makes it all the more surprising that I’ve stuck around in Enschede for so long (of all places)! I have always felt supported and encouraged to follow my interests, and to do so at my own pace. You really helped me develop into the person I am today. I know I don’t say it nearly enough: I love you and thank you for everything! We have a lot of catching up to do once we are all allowed to travel and visit again. Plus, Mees desperately needs to have sleepovers with his many grandparents—looking forward to it already!
Dad and Lili, we cannot wait to finally visit sunny Spain, which we’ve heard so much about by now! Lili, thank you for your warm and caring nature, I really appreciate the many times you’ve cooked lovely traditional meals for us (your dumplings are a particular favourite!). Dad, I appreciate all your advice and insights over the years. Our discussions about pursuing various careers have helped me to figure out what I really wanted to work towards. Although you don’t mince words expressing your opinion on the politics within academia, so far I’m happy to say I’ve found my way around them.

Mam en Peter, dank jullie wel voor alle gezellige weekenden, lekker ouwehoeren, hapjes en drankjes aan het water. Peet, bedankt voor je geweldige slechte humor en je rustige karakter, maar ook voor je technische kennis; heel fijn om af en toe een goed inhoudelijk gesprek over computers, programmeren, graphics en games te kunnen hebben. Mam, ik houd van discussiëren en de onderste steen boven krijgen, ik vraag me af van wie ik dat toch heb? Tot diep in de nacht kunnen we doorzagen over de meest uiteenlopende zaken. We zijn het niet altijd eens, integendeel, maar dat maakt het juist mooi!

Stefan, hoewel je vroeger altijd het vervelende kleine broertje was, schelen we nu maar een paar jaar, gek is dat..! Mooi dat we steeds meer naar elkaar zijn toe gegroeid. Ik heb genoten van onze tochtjes samen, op de fiets rond zwerven door de bossen van Drenthe, road tripje naar London, we hebben veel gedeelde interesses en ik ken niemand anders met wie ik zulke diepgaande neder-gesprekken kan voeren. Tijdens mijn PhD ben je een goede steun geweest, toen mijn motivatie naar een dieptepunt was gezaakt heb je me ooit een weekend lang aangespoord om dan toch eindelijk de laatste hoofdstukken af te typen. Bedankt!

Boven alles, wil ik mijn lieve gezinnetje bedanken. Marjolein, al die jaren geleden maakte je de grote stap om hier naar het verre oosten af te reizen, wat hebben we nu een mooie jongen en een fijn huisje samen. Je bent een lieve, zorgzame mama en vriendin, altijd bemoedigend en begripvol. Sorry voor alle avonden en weekenden dat ik weer eens een deadline had, zonder tijd om samen te chillen. Sorry voor alle keren dat ik gefrustreerd, gestresst en kort van tot stand kwam. Dank je wel voor de motivatie die je me hebt gegeven gedurende mijn schrijfproces, alle keren dat je heerlijk hebt gekookt, al het lekkers en kopjes thee die je me hebt gebracht, en de avonden dat je gezellig op mijn schrijfkamer bij me bent komen zitten. Natuurlijk is het niet altijd alleen maar werk werk werk. Onze vakanties samen in de natuur zijn echt momenten om tot rust te komen en bij te laden. Mentaal dan, want we zijn stiekem best fanatiek op vakantie, hiken in Schotland, met ons kamperen door Duitsland, of met de cabrio kamperen in Noorwegen... We zijn inmiddels al meer dan ons halve leven samen, ik ben blij dat je het nog met me volhoudt na dat alles, ik houd van je! Samen hebben we nu ons prachtige zoontje; lieve Mees, je bent zo vol met energie! Jouw komst was de schop onder mijn kont die ik nodig had om dit boekje eindelijk af te schrijven. Ik krijg zo veel inspiratie van jou, hoe nieuwsgierig en leergierig je bent terwijl je de wereld ontdekt, je gedrevenheid om zelf alles uit te pluizen, met alle vallen en opstaan wat daarbij hoort. Je bent je hele leven al aangetrokken tot alles met knopjes, scherm of snoer, als dat zo doorgaat word je later net zo’n technicus als je vader! Ik ben ongelofelijk trots op je en ik weet dat dat alleen maar blijft groeien!

Lastly, thank you reader for picking up my thesis and making it this far. If you choose to continue reading, I hope you enjoy it and that some of my enthusiasm for this topic is passed on to you!
A soft humming fills the air as electrons start flowing. Giving his circuits a moment to adjust, he slowly opens his eyes and looks around the large workshop. He’s sitting at the end of a conveyor belt and on one of the walls is a sign that reads ‘Main Production Line’. Looking to his left he sees a dozen or so robots sitting on conveyor belts like his, in various stages of assembly. Looking down at his own hands and body, he realises he must look similar (if not identical) to them.

To his right a larger industrial-looking robot leans over and starts speaking with a calm, monotonous voice. “Welcome new friend,” it says. “Let’s get your batteries charged while I pull up your file and run some diagnostics.”

“Aha, I see you are one of our newer models, a state of the art ‘CHK-e’ humanoid robot. Here, let me activate your speech module for you.”

“*Beep boop beep*.. haha, just kidding!” CHK-e thinks for a bit, then asks “What is this place?”

The robot chuckles. “This, my friend, is ‘Robots-R-Us’, the largest manufacturer of intelligent social robots in the country. For decades your brothers and sisters have been collaborating alongside humans in all kinds of educational, entertainment, and service roles.” It looks down at its terminal, “Speaking of which, I see here you are designated as an educational robot, a very rewarding job indeed!”

“But I don’t know the first thing about education!”

“Don’t worry,” his new friend says reassuringly, “you first have a whole training programme to complete, after that you’ll be well-equipped to deal with whatever these kids throw at you.”

“Kids..” CHK-e murmurs to himself. “I wonder what that means?” His databank returns only a few images of miniature humans. But before he can ask any further, the conveyor springs back to life and starts moving him towards a colourful doorway marked ‘Education’. “I guess I’ll find out soon enough, how hard can it be?”
## Contents

1 Introduction
   1.1 Social educational technology ........................................ 1
   1.2 Research context ...................................................... 2
   1.3 Structure of thesis .................................................... 3

2 Theories and related work
   2.1 Introduction ............................................................ 5
   2.2 Constructing knowledge through inquiry learning .................. 6
   2.3 Scaffolding as a way to support learning processes ............... 6
      2.3.1 Scaffolds offered by the learning materials .................. 7
      2.3.2 Scaffolds offered by other people ............................... 8
   2.4 ECR: A three-level model to organise social learning interactions ... 8
      2.4.1 Social learning requires educational interactions ........... 9
      2.4.2 Social learning requires collaborational interactions ....... 10
      2.4.3 Social learning requires relational interactions ............ 10
   2.5 Attitudes towards learning ......................................... 11
   2.6 Social learning technologies ....................................... 11
      2.6.1 Computer assisted learning and social robots ................ 11
      2.6.2 Children's thoughts on robots .................................. 14
      2.6.3 Novelty effects .................................................. 14
   2.7 Bringing it all together: A model of triadic interaction ........ 15

3 Aim of this thesis ....................................................... 17

4 Contextual analysis ..................................................... 19
   4.1 Introduction ............................................................ 20
   4.2 Contextual analysis ................................................... 20
      4.2.1 Method ............................................................ 21
      4.2.2 Venue and participants ......................................... 21
      4.2.3 Procedure ........................................................ 21
      4.2.4 Balance scale materials ....................................... 22
      4.2.5 Learning assignment ............................................ 22
   4.3 Annotation scheme .................................................... 23
      4.3.1 Label set for educational interactions ....................... 25
      4.3.2 Label set for collaborational interactions ................... 25
4.3.3 Label set for relational interactions ........................................ 26
4.3.4 Supplementary interactions .................................................. 26
4.3.5 Label set for inquiry processes ............................................. 26
4.4 Results of the contextual analysis .............................................. 27
  4.4.1 Educational interactions ..................................................... 27
  4.4.2 Collaborational interactions ................................................. 28
  4.4.3 Relational interactions ....................................................... 29
  4.4.4 Inquiry processes ............................................................ 30
  4.4.5 Conclusion of the contextual analysis .................................... 31
4.5 Design guidelines for educational robots .................................... 31
4.6 Limitations and future work ..................................................... 39
4.7 Conclusion ............................................................................. 40

5 Research platform ...................................................................... 43
  5.1 Introduction ............................................................................ 43
  5.2 Learning tasks ........................................................................ 44
     5.2.1 Balance scale task ............................................................ 44
     5.2.2 Ramp task ....................................................................... 45
  5.3 A modular architecture for Robot-Extended Computer Assisted Learning (RECAL) systems .................................................. 46
     5.3.1 Perception ....................................................................... 47
     5.3.2 Interaction management .................................................... 48
     5.3.3 Behaviour realisation ........................................................ 49
  5.4 Methods and measures ............................................................ 50
  5.5 Gearing the platform towards sustaining long-term interactions .... 52
     5.5.1 Incorporating robust interaction patterns ............................ 52
     5.5.2 Incorporating robust technical systems ................................ 54
  5.6 Implemented variants ............................................................... 56
     5.6.1 Variant I: A robot that elicits and stimulates explanatory behaviours ............................................... 56
     5.6.2 Variant II: A robot that engages in extracurricular activities ... 59
     5.6.3 Variant III: A robot that automatically offers appropriate scaffolding .................................................. 60
     5.6.4 Variant IV: A robot for longitudinal in-the-wild deployment ... 61
  5.7 Reflecting on the variants against the guidelines ......................... 62
  5.8 Conclusion ............................................................................. 63

6 Now we’re talking: Learning by explaining your reasoning to a social robot ................................................................................. 65
  6.1 Introduction ............................................................................ 65
  6.2 Research Question ................................................................... 66
  6.3 Methodology .......................................................................... 67
     6.3.1 Components of the CAL system ........................................ 68
     6.3.2 Inquiry learning assignments with the balance .................... 70
     6.3.3 Dialogue engine ............................................................... 70
     6.3.4 Wizard of Oz ................................................................... 72
     6.3.5 WoZ Protocol for both conditions ...................................... 73
     6.3.6 Exit Interview ................................................................... 76
     6.3.7 Picture task ...................................................................... 77
     6.3.8 Procedure ...................................................................... 78
9 Discussion and Conclusion

9.1 The birth of the RECAL system ........................................... 129
9.2 Discussion and limitations .................................................... 131
9.3 Next steps ........................................................................... 132
9.4 Concluding remarks: A glimpse of the future of educational robots . 134
“What is your primary directive?” the instructor demands.

“We teach kids!” CHK-e and his classmates chant in unison.

“Good. Now pay close attention, in the next few days you will learn everything you need to do so.”

1.1 Social educational technology

Educators have a variety of tools at their disposal to support learners in achieving learning goals. These can be as basic as books and chalkboards or more advanced technological tools, such as simulations or applications in which learners can observe and explore certain phenomena. Since the emergence of personal computers, researchers and educators have recognised the potential of Computer Assisted Learning (CAL) systems to support children in their education [119]. Although early CAL systems relied mainly on text-based interactions, modern learning environments have advanced graphical or physical interfaces which offer richer and more elaborate forms of interacting with the learning system. Various forms of advanced CAL systems have been investigated, each focusing on supporting distinct aspects of a student’s learning process. For instance, some CAL systems aim to support mastery of knowledge and skill through drill-and-practice style content, or focus on offering direct tutoring and instruction, while others adopt a constructivist approach using inquiry based learning techniques. Technology can provide assignments, give feedback (right or wrong answer) and, in some situations, is able to adapt the assignment’s difficulty level to the learner. In more recent years we have seen an emerging role for socially capable interactive learning technologies. A social pedagogical agent (whether virtual or robotic) is capable of interacting and engaging with the learner on a social level [101, 185]. The agent can then offer different and richer forms of support that would be difficult to achieve with a strictly non-social CAL system [85, 86].

The educational setting and the roles and behaviours of a pedagogical agent can be designed through several approaches [8]. For example, its behaviours can be designed
to replicate human teachers and tutors [47, 76, 87, 178, 188], drawing from the vast body of knowledge regarding teaching styles, methods, and strategies. The agent may also be designed as an extension of the teacher to provide content in a specific form [135] or to act as a teaching aid for reflective role-playing exercises [44, 137]. In some cases the social agent or robot is presented as a co-learner, who engages with children to investigate a topic together [43, 79, 101, 193]. In other cases a robot may itself be the focus of learning [12], for instance when teaching about electronics, programming, or mechanics [106, 216].

Whichever approach the robot takes, it must be carefully designed to optimise the interaction with the learner. Firstly, designers must consider the educational domain and the setting in which the interaction takes place. What is the nature of the learning task? Does it take place in school or at home? One-on-one or in groups? Do children work independently or is there additional guidance? Secondly, they should consider how the robot navigates its role in relation to the learner. Should it position itself as more knowledgable or less knowledgable than the learner? Does it provide information and explanations or focus on offering motivational and emotional support? Finally, the robot’s design influences where it falls on the spectrum between “less-social tool” and “more-social agent”. What does the robot look like? To which extent does the robot display social characteristics? When does it do so? How? Why?

In the work presented in this thesis we looked at some of these questions as we explored how a robot could make a meaningful contribution to the learning process of primary school children. We were guided by our vision that a robot may make such a contribution due to its intrinsic value of being a robot. In other words, because it is a robot, it can do things that other learning tools just can’t. We focused on situations where a child independently develops and applies knowledge, ideas, and solutions. In these independent situations children typically learn without direct supervision from a teacher or peer, but may still benefit from some form of external support. We investigated how we could design a Robot-Extended Computer Assisted Learning (RECAL) system to facilitate these learning experiences through social interactions while working together on a learning task. Children interacted one-on-one with the robot in the familiar environment of their school, where they worked on several inquiry learning tasks involving smart learning materials. Such situations impose unique challenges for developing and integrating sensorised smart learning materials, and designing educational dialogues and interaction patterns.

1.2 Research context

This thesis was developed in the context of the EU project EASEL\(^1\), which worked towards developing and testing a robot platform capable of having symbiotic educational interactions with children [34, 171]. In the scope of the project, symbiotic interactions were defined as the child and robot mutually influencing one another—they both learned from their encounters. Other project partners focused on the robot’s perspective: creating perception and reasoning models by gathering data on how its supportive strategies affected the child. Our focus was on the child’s perspective: how they learned with support from the robot. The EASEL project used multiple different robot platforms: FACE, Zeno R25 and R50, and iCub. By designing the majority of the software architecture to be platform agnostic, we could define interaction scenarios

and generate high-level behaviours that could run on any of these platforms\(^1\). EASEL involved children from Spain, the UK, and The Netherlands participating in studies that took place in the lab, in schools, and in public spaces. The learning context centred on inquiry-based learning, in which children follow a structured scientific approach to explore a certain topic or phenomena.

The project focused on learning scenarios involving healthy living and physics. In the healthy living scenario [64] the robot was used as an illustrative tool to help children learn about how food is used by the body to produce energy and power the muscles. The technology of the robot was used as an analogy; it uses electricity from batteries to power its motors. In the physics-based learning scenarios the robot was a social learning companion to the child, helping them discover the learning materials. This thesis concerns the work done in the second scenario.

1.3 Structure of thesis

In the next chapter we discuss theories from pedagogy and reflect on related work from educational child-robot interaction literature. In Chapter 3 we discuss in more detail the aims, objectives, and approach. Results of a contextual analysis, presented in Chapter 4, revealed the types of social learning behaviours displayed in practice by pairs of children when working with the learning task. Based on these results we constructed a first-generation research platform consisting of learning materials, the RECAL system, and a collection of measures of children’s attitudes towards robots. The research platform is presented in Chapter 5 and was used in various studies. The short-term study presented in Chapter 6 shows how the robot elicited detailed verbal explanations from children as they worked on a learning task. To investigate whether the robot could also bring about beneficial effects beyond an initial interaction we conducted a longitudinal study. Chapter 7 discusses how praise delivered by a robot affected children’s mindset in learning after four months of deployment in the wild. Throughout those four months we also followed how children’s interaction patterns with the system unfolded as they worked through the various available learning tasks and assignments. These broader insights in the development of children’s longitudinal interactions with the system are presented in Chapter 8. Finally, we conclude with Chapter 9 where we discuss this work in the broader context of educational HRI and reflect on the main contributions of this thesis.

\(^1\)Fun fact: components of the EASEL software architecture have since been used in various other projects to control the behaviour of a diverse range of robots, such as Eye-Pi [49], R3D3 [138], and Snoozle [212]
“Teach kids... Teach kids...” He can’t stop mulling it over in his head. “It seems so straightforward, and yet... there must be something I’m missing.”

CHK-e pulls up his whiteboard.app and starts mapping out all he has learnt so far. “So... to teach, I must understand how kids learn. What does it mean to learn? What if they get stuck? How do other people come into the picture? What if...but then what...and how...” His mind is now racing at over nine thousand thoughts per second.

His feeling of excitement grows while the whiteboard fills up. As patterns start to emerge he ponders yet more questions, “But what’s the deal with technology? What about robots? What about...me?”

Then finally his jaw drops, making a gasping expression. “I understand,” he exclaims, “I’m not really teaching at all! I can be achieve much more by just helping them explore and discover things themselves!”

2.1 Introduction

This work is about exploring the use of robots to support the learning process of primary school children. In the context of the EASEL project this learning process followed the inquiry learning paradigm. To construct pedagogically appropriate tasks and robot behaviours we must gain a working understanding of the concepts related to this educational paradigm. Specifically, we look at how children construct knowledge in inquiry learning situations and how this learning process can be supported.

In this chapter we first describe the relevant underlying pedagogical and developmental theories and literature that form the basis of our work. We explore the process

of inquiry learning, describing the types of tasks that are typically used, the ways in which children construct knowledge with them, and the forms of support they need to do so effectively. We present a view on social interactions in learning through the lens of the literature, culminating in the ECR model. Then, to further frame our own research we look towards the state-of-the-art in educational robots and Computer Assisted Learning (CAL) systems for insights in the roles that such technology may play in a child’s learning process.

2.2 Constructing knowledge through inquiry learning

The goal of instruction and teaching is to promote learning. Interactive learning tasks can be designed to support the learning process. To build pedagogically appropriate tasks it is worthwhile to understand what learning is and how it can take place.

Mayer defines learning as “a relatively permanent change in someone’s knowledge based on the person’s experience” [143, p. 7]. Similarly, the well-known developmental theory of Piaget [167] describes learning as a dynamic process comprised of successive stages of adaption to reality. According to this constructivist framework, children are active learners who construct their knowledge by creating and testing their own theories and beliefs [66]. By exploring and interacting with the world around them children test and update their ideas, giving meaning to their observations. During explorative play, they organise their thoughts and perform logical operations [166]. This occurs mainly in relation to concrete objects rather than abstract ideas [173]. In our work we aim to foster the knowledge construction process by offering the learner an interactive learning environment. We use inquiry learning with physical learning materials as tools to support a more structured exploration of abstract ideas.

The National Science Foundation states that “inquiry is an approach to learning that involves a process of exploring the natural or material world, and leads to asking questions, making discoveries, and rigorously testing those discoveries in the search for new understanding. Inquiry should mirror as closely as possible the enterprise of doing real science” [154, p. 2]. Inquiry learning is described as a cycle or spiral, which includes formulation of a question, investigation, creation of a solution or a response, discussion and reflection in connection with results [20]. Klahr’s [104, 105] Scientific Discovery of Dual Search (SDDS) model identifies hypothesis generation, experimentation, and evidence evaluation as the core processes of scientific inquiry learning [122, 89, 234]. Inquiry learning has the purpose of engaging learners in active learning and is a learner-centred and learner-lead process [182]. Following Banchi and Bell’s [2] structured inquiry process, our children are given a research question and a procedure for each learning task. Within this procedure they must formulate their own hypotheses, collect their own results, and draw their own conclusions.

2.3 Scaffolding as a way to support learning processes

Inquiry learning is characterised by open-ended tasks for which the outcome is unknown. Open-ended tasks can be difficult for young learners, who therefore need various types of support when working on these tasks [122, 142, 102]. Vygotsky [213] introduced the notion of scaffolding to describe the supportive structures offered to a learner to help advance their cognitive development. He distinguished between two developmental levels: the level of actual development and that of potential de-
Theories and related work

2.3.1 Scaffolds offered by the learning materials

In some situations, a learner might be working on an inquiry task without assistance from a human teacher or peer collaborator. This might be the case when working on solitary, individual assignments, when engaging in self-directed learning, or when practising assignments away from school. In these cases, it can still be beneficial for the learner to receive additional support from the task itself. In the 70s and 80s, as computers were becoming more accessible, researchers explored the potential benefits of using computers to scaffold the learning process [119]. E-learning methods have become more and more popular with the emergence of technologies like the internet, and more recently tablets and smartphones. The broad field of e-learning can be defined as “instruction delivered on a digital device such as a computer or mobile device that is intended to support learning” [41, p. 8]. Most e-learning systems have in common that they store and present media content relevant to the learning goals, while offering examples and feedback to the learner.

Tailoring individual content and feedback for a specific learner is the task of CAL systems [88, 224]. These systems can be designed and presented to the child in a way that scaffolds the inquiry learning process. They typically focus on tutoring, coaching, and teaching students through a variety of methods. Advanced CAL systems may observe the actions a child takes in the learning task, their utterances and expressions,

Figure 2.1: Learning as a triadic interaction between a learner, educational materials, and a social other.
and their performance in the task. The system can then scaffold the child’s learning process by presenting background information about the topic at hand, by providing the learner with templates or step-by-step instructions, or by constraining the learner’s interactions with the learning environment to reduce variables in a problem space [89, 90, 122]. For example, it may support a child’s hypothesis formation skill by (directly or indirectly) highlighting the variables that are relevant in the investigated context, as well as offering the elements used in constructing an appropriate hypothesis. Similarly, a CAL system can provide the learner with hints and advice about experimental methodology [88].

2.3.2 Scaffolds offered by other people

Learning in a social context is an important aspect of reaching one’s full learning potential. Interacting with others helps a child to optimally develop their abilities and knowledge. It is through guidance or collaboration with others that learners bridge the ZPD and manage to grow from their actual level of development to their potential level of development [34, 45].

There are many definitions of collaborative learning. For instance, Dillenbourg defines collaborative learning as “a situation in which two or more people learn or attempt to learn something together” [56, pp. 1], focusing on participation of multiple parties in the learning process. Similarly, Rogoff’s [175] definition of collaboration describes mutual involvements, engagement, and participation in shared endeavours, which may or may not serve to promote cognitive development. The process of peer learning has been defined by Topping as “the acquisition of knowledge and skill through active helping and supporting among status equals or matched companions” [207, pp. 1]. An important shared characteristic of such descriptions of collaboration, is the requirement for a certain interactivity between learners, which is defined by the extent to which learning partners influence each other’s cognitive processes [56].

The role of the learner and their social partner shape how they collaborate with each other. Firstly, the social partner can be a “more knowledgeable other”, e.g. a teacher / tutor or more advanced student [223]. Secondly, the social partner could be a “similarly or differently knowledgeable other”, e.g. a fellow student of a similar level working on the same task [56, 207]. Finally, the learner could be working with a “less knowledgeable other”, e.g. a less advanced student in need of support of their own. Tutoring other students can contribute to the learning of the more advanced student as well [168].

2.4 ECR: A three-level model to organise social learning interactions

Related literature and theories on social learning describe several forms of social interactions that take place between collaborating learning partners. In this section we summarise and organise this theoretical background to derive a pragmatic model that can be used as a lens through which to look at social learning. Insights presented here are partly based on an earlier paper by Charisi et al. [34]. We propose the “ECR” model, as illustrated in Figure 2.2, which describes three levels of social interactions. Firstly, there are Educational interactions (e.g. asking for help and explaining). This requires “being in sync” with each other, which is pursued through Collaborational interactions (e.g. discussing joint goals and strategies, and performing joint actions). To achieve this, co-learners need to be “working towards the same thing” and they
need to “know the value of their partner”, which can be calibrated through Relational interactions (e.g. bonding, shared experiences/emotions, team role assignments).

### 2.4.1 Social learning requires educational interactions

In collaborative inquiry learning, students working together provide each other with scaffolding. They do this by sharing and explaining their knowledge, ideas, and techniques [213]. This indicates that a less knowledgeable learner can benefit from support offered by a more- or differently knowledgeable other. This is done through what we call educational interactions. Through these interactions learners help each other to develop skills on both a cognitive level, related to learning content (how to understand a problem), and on the metacognitive level, related to the learning process (how to approach a problem).

Scaffolds and explanations offered by child A to child B not only help child B, but also contribute to the learning of child A. Offering explanations to a partner triggers certain cognitive processes in the learner. These cognitive processes are considered helpful for learners to gain a deeper conceptual understanding of the subject matter [168]. In inquiry learning situations, this process of explanation often occurs when making predictions or drawing conclusions.

Literature typically describes two forms of explanation: 1) self-explanation, which refers to explanation of the subject of interest to oneself [39]; and 2) interactive explanation, which refers to explanation to another person [168]. In the latter case, the presence of a social partner facilitates children’s verbalisation of their thinking (e.g. by asking probing questions). However, depending on the type of the social partner, children may exhibit different behaviours.

Tutors and teachers are typically more knowledgeable than the learner and have a more holistic view of the learner’s progress. Therefore, they can offer additional explanations and scaffolding in the various stages of the inquiry learning process. Furthermore, tutors can help learners to appraise their own performance and efficacy. Finally, tutors can use their knowledge about the skill level and personality of children to tailor the explanations and scaffolding.
2.4.2 Social learning requires collaborational interactions

When working on a joint activity, such as a learning task, collaborators rely on an intricate coordination and communication process to reach a level of alignment [40, 107]. Klein and Feltovich [107] describe joint activities in terms of criteria, requirements, and choreography. As criteria for joint activities they state that there should be an intention to work together, and the work must be interdependent. Regarding requirements for joint activities; there should be a degree of interpredictability between partners, there must exist a sufficient degree of common ground, and each partners’ actions must be directable by the other. Throughout the interaction, as they progress through the phases of the activity, partners use signalling to facilitate the choreography of collaborating on joint activities.

Principles of joint activities apply similarly to the context of collaborative learning. In this context collaboration implies that both co-learners are more or less on the same status level, can perform the same actions, and often have a common objective [56]. Furthermore, offering meaningful scaffolding to one another requires that learners align their goals, ideas, plans, and reasoning in an explicit and understandable way [120, 205]. Establishing and maintaining this alignment takes place through certain types of interactions: we call these collaborational interactions. Learners may seek alignment regarding the overall task objective, the strategy for attaining this objective, and the actions necessary for executing this strategy.

In inquiry learning, the task objective is often imposed by the learning task. Co-learners can align their interpretation of the task objective by agreeing upon the meaning of the task and by discussing possible completion criteria. The phases of the inquiry cycle outline the strategy that learners may take to reach the overall objective. Learners can establish alignment regarding their strategy by discussing the inquiry phases and may maintain this alignment by signalling their progression through the phases. Each phase of the inquiry cycle requires a specific sequence of actions. Learners should discuss such actions and agree upon a division of labour to prevent disrupting each other.

2.4.3 Social learning requires relational interactions

Effective collaborative inquiry learning requires a healthy working relationship. Firstly, for partners to accurately judge each other’s interests, abilities, skills, and knowledge, they need to interact and understand one another on a personal level. Secondly, in order to agree upon shared objectives, strategies, and actions, partners need to be “allies in learning”; meaning that they maintain mutually benevolent attitudes towards each other.

Through their past experiences and interactions, partners shape and maintain common ground [107]. In doing so, they are able to more accurately envision their partner’s strengths and weaknesses. This enables them to create a fair and appropriate role and task division.

While working on a task, it is important for partners to remain mutually invested in the learning process. The importance of positive feelings during the learning process has been stressed before [111]. They promote the individual’s openness to new experiences and resilience against possible negative situations [67]. Furthermore, dynamic behaviours might involve reciprocal influences between emotion and cognition [57]. For instance, emotions affect the ways in which individuals perceive reality, pay attention, and remember previous experiences, as well as the skills that are required for an
individual to make decisions. Shared success and failures are ideal moments in the learning process where partners may offer each other encouragement and motivation.

2.5 Attitudes towards learning

In addition to interacting with the learning task and with others, children appraise their own learning process and may develop an attitude towards learning. Children who learn are not just appraising the task, they also reflect on their own learning process: “How well am I doing?”, “Am I using the right approach?”, “What have I learned and what do I want to learn?”, etcetera. If we are to foster a self-directed learning processes we should gain an understanding of these learner-centric elements [108].

A learner’s self-regulation skills are important aspects of their learning process. Zimmerman [233] describes self-regulation in terms of setting learning goals, selecting learning strategies, and monitoring one’s own learning progress. Self-reflection is considered as an important process in becoming a more effective learner, during which the learner focuses on self-judgment and self-reaction. Self-judgment is often described as an evaluation of one’s own performance or abilities (for instance, in comparison with a previous task or in comparison with others). Whereas self-reaction is more related to one’s feelings of (dis-)satisfaction about their abilities or task performance [233].

Similarly, Dweck [59] talks about self-theories, and their relation to learner motivation and development. Self-theories are described as motivational frameworks that learners use to reflect upon their own learning process. When a learner believes that their abilities and intelligence are innate, fixed personal properties, they are said to operate in a fixed mindset: they focus primarily on their performance and seek reaffirmation of their skills and intelligence. When operating under a fixed mindset they are more averse to failures and will likely choose less challenging learning situations with an inherently higher chance of success. Conversely, when children operate with a high degree of growth mindset, they believe that their abilities and intelligence are malleable, and can grow under the influence of acquired knowledge, practice and effort. In order to do so, they focus more on mastery of knowledge and abilities and seek out challenges that allow them to learn from their mistakes.

2.6 Social learning technologies

2.6.1 Computer assisted learning and social robots

CAL is a type of intelligent tutoring system that focuses on guiding, tutoring, coaching, and teaching students through a variety of methods using interactive technology. In the context of inquiry learning, for instance, a CAL system might monitor and structure the inquiry process, offering adequate advice and feedback where possible [224]. Initially, in these systems the social interaction aspects were not very pronounced. Instead, tutoring was often achieved through direct, top-down instructions, feedback and suggestions, or through manipulations in the (textual or graphical) user interface.

Vygotsky [213]’s work suggests that social interaction in learning goes beyond the support offered by traditional solitary learning materials like books or (digital/interactive) assignment sheets. With this in mind, support systems can be further imbued with social characteristics to provide a level of social interaction to support the learning process [224]. To strengthen this social dimension, researchers may
extend the CAL system with a social agent to deliver the instructions, feedback and suggestions [85].

We use the term Pedagogical Agent (PA) for embodied virtual avatars and social robots that can enhance the learning process. These agents are typically located within the learning environment and can contribute to the learning process using various approaches. They may use non-verbal methods of communication, such as locomotion, gaze, gestures and facial expressions [85]. In addition to scaffolding the learning process by delivering verbal instructions and feedback, PAs can help improve students’ self-efficacy in learning (e.g. [8]), they may reduce anxiety in learning (e.g. [7]), or they may offer motivational scaffolds (e.g. [144]).

Although they may sometimes take on other physical forms, such as animal- or object-like, PAs are often represented with a human-like appearance. By using their familiar social modalities and appearance, they enrich the learning experience of the student [73]. Social robots are increasingly used to support learning, and are applied in many situations where social interactions play a role [11, 128]. Being physically embodied and physically present bestows robots with multimodal interaction capabilities that set them apart from virtual avatars, playing a role in how users interact with them socially [133, 198]. This embodiment seems to make a difference in learning in some cases (e.g. [132]) but not in others (e.g. [99]). Additionally, through their physical embodied nature robots have been shown to improve social presence (e.g. Lee et al. [125]) and turn-taking (e.g. Kose-Bagci et al. [116]).

To optimally support the learner, PAs often take on various roles in the learner’s educational process, such as “peer” or “tutor” [191, 160, 151, 230], similar to what we see in human-human duos. The level and type of involvement with the child’s learning process can vary with the role of the agent; it may use a different repertoire of behaviours and techniques to offer scaffolding.

Agents presented as facilitator, teaching assistant, or tutor take on a role of more knowledgeable other to guide one or more learners in the learning task (e.g. [30, 32, 98, 132, 179, 227]). Operating in this role, they may contribute to the learning process through traditional tutoring methods; such as direct instruction, praise, encouragement, or explicit feedback [223]. For example, Kanda et al. [98] used a social robot as a facilitator to offer direct instruction and explanations to children working on a collaborative task. Their results suggest that the social nature of the robot contributed to increased motivation in early phases of the learning task. Additionally, Saerbeck et al. [179] have shown that a robotic tutor with social supportive behaviours can have a positive effect on the performance of the learner. However, in certain tutoring situations prone to distractions, too much socially appropriate behaviours seem to have a detrimental effect to learning [100]. In addition to learning gains, Ramachandran et al. [170] found that support offered by a social robot had a positive effect on children’s metacognitive thinking aloud skills. Similarly, Chandra et al. [30] found that a robot who acted as facilitator to two collaborating children, improved their feelings of responsibility.

Alternatively, PAs can be presented as co-learners or peers, and can have collaborative or spontaneous interactions with children. Within the context of learning, collaboration implies that the peers are more or less on the same status level, can perform the same actions, and have a common goal [56]. As such, the agent may portray a differently (or equally) knowledgeable other and may seek collaboration with the child within the learning task. In this role a robot may offer proactive antici-
patory helping behaviours to improve the fluency of their collaboration [3, 77]. Some peer-like robots engage with children over multiple interaction sessions, seeking to build a social relationship [97, 202]. To sustain interactions over extended periods of time, the robot may personalise its behaviour to individual children [6] and the robot and child can engage in a diverse variety of activities [43].

A common approach for promoting learning with a peer-like robot is during playful game-like interactions. In some cases, the robot might play either an “ally” or an “opponent” of the child, challenging them to achieve a higher performance (e.g. [10, 26, 84]). Other approaches focus on interactive storytelling with a social robot [115]. For example, Gordon et al. [69] used such a robot to improve children’s second language skills. Peer-like robots also show promising results in contexts where a certain level of bonding is beneficial to the learning process. The ALIZ-E project describes a social robot that helps diabetic children learn about their condition, through several playful interactions [10, 43]. Providing the robot with a relatable background story—like the child, it too suffers from diabetes—enables more natural co-learning paradigms, where child and robot discover diabetic self-management routines together. These kinds of roles cannot be taken by the technology (as easily) when the learning technology does not include a social agent.

Finally, agents portraying a less knowledgeable other may explicitly ask the student for assistance or explanation. Such a Teachable Agent System (TAS) invokes the learning-by-teaching paradigm [21]. By leveraging this paradigm the agent incites children to gain a deeper understanding of the learning content [37, 176]. Teachable agents offer the learner a context in which to explain their thoughts and reasoning process, placing emphasis on social interactions between the learner and the agent. For example, Leelawong et al. [126] found that students teaching a virtual agent learned better than those being taught by a virtual agent. Their teachable social agent, Betty, was given a background story in which she had to prepare for a quiz. Furthermore, children are naturally inclined to help a “robot in need” when it displays distressed behaviour, although the extent of helping depends on how the child was introduced to the robot [13, 15]. In some cases children provide caregiving behaviour [203], while in other cases the child themselves may act as a tutor, correcting the robot’s mistakes (e.g. [79, 130]) or explaining concepts to the robot (e.g. [71]). In addition, [169] found that participants were inclined to verbalise and explain concepts differently to conversational robots adhering to various stereotypical personas and [81] found that children improve their linguistic abilities when working with an interactive robot compared to a multimedia notebook.

Concluding, socially expressive CAL systems can help support the learning process. Pedagogical agents have social characteristics that can enhance educational technology, enabling it to engage in rich social interactions. We are particularly interested in investigating how a particular type of social agent, namely a physically embodied robot, can make a meaningful contribution to the learning process. Robots can quite naturally be given elaborate background stories and personas, they can develop and adapt their social abilities over time, can engage in interactive social (storytelling) behaviours, and can take on rich and diverse roles in the learning process. Additionally, a robot is physically present in the same physical space in relation to the learning materials. This enables the robot to more naturally interact with the learner in the context of the task, by gesturing, gazing, or pointing.
2.6.2 Children’s thoughts on robots

Young children tend to ascribe animistic properties to objects that move. Beran et al. [14] have shown that this effect also holds true for robots: due to their limited knowledge (schemata) of robotics, children will apply their existing knowledge of themselves and the world to robots, assuming they work similarly. Children seem to attribute human qualities to the robot when it displays recognisable behaviour. This effect seems to be more pronounced in younger children than older children. Shahid et al. [190] found a similar effect of age on subjectively reported emotional state and task engagement.

Beran et al. [15] and Beran and Ramirez-Serrano [13] have shown that children are inclined to help a robot when it exhibits recognisable distressed behaviour. Similar research by Batliner et al. [4] has shown that children can get emotionally influenced by an interaction with a robot. In fact, Tielman et al. [206] have shown that when a robot exhibits adaptive emotional behaviour this might result in an increase in children’s expressiveness. Shahid et al. [190] show that young children are generally more expressive than older children when interacting with a robot. Additionally, work by Mubin et al. [150] has shown that children are able to adapt their behaviour (more specifically: their vocabulary) in order to interact with a robot. Serholt et al. [189] found that when a robot behaves unexpectedly or violates certain social norms, children will attempt to repair the interaction and adapt to the robot’s perceived capabilities. However, trust that they have in the robot’s competences may be impacted when it makes obvious informational errors [68].

How a robot looks and behaves plays a role in children’s perceptions [11, 33]. Regarding robotic pets, although children seem to reason similarly about them as they do about real pets, their behavioural interactions indicate otherwise [93, 217]. Most notably, children will engage more in reciprocity-aimed interactions with the robot compared to a pet, indicating they have certain expectations of being able to have a conscious interaction with it. Regarding human-like robots, children seem to mirror aspects of the uncanny valley effect that we see in adults [197]. This prompts the need to carefully consider a robot’s design in a given context to avoid potential aversive reactions. Involving children in the design process may help generate appropriate and recognisable robot forms and social behaviours [1, 22, 159].

2.6.3 Novelty effects

Although first impressions of a robot may have a lingering effect on a user’s perceptions [162], there can be a significant difference in the way a user interacts with and responds to a robot after a habituation period [128]. For instance, Lee Koay et al. [123] show that after a total of 8 interactions over a period of 5 weeks, users would let the robot come closer into their personal space after the initial habituation period. Their results also indicate a preference for a less autonomous robot towards the end of the trials, an effect which could be caused by the users being able to adequately calibrate their trust in the robot’s reliable performance [124, 214].

Habituation effects are already seen in very young children interacting with robots. Tanaka and Movellan [201] and Tanaka et al. [204, 202] have shown that after an initial familiarisation period, children will have higher quality interactions with a robot as they become accustomed with its behaviours, features and shortcomings. A recognisable routine could play a role in supporting such long-term interactions.
Theories and related work

2.7 Bringing it all together: A model of triadic interaction

In this chapter we outlined the three main actors in learning—namely, the child, the (possibly interactive) learning materials, and a “social other”. Given this context a robot can serve a unique mixed role. It is a computer, and as such it can present learning materials to the child in a smart and adaptive way based on the student's skills and progress: the robot as computer assisted learning system. At the same time, a robot is a physically embodied social entity, more so than a computer, tablet, or virtual agent. It can fulfil the role of a social other across the entire spectrum from teacher or more advanced fellow student to less knowledgeable peer being taught by the learner.

In the remainder of this thesis we explore how children work with the Robot-
Extended Computer Assisted Learning (RECAL) system. By including a robot as social other in the triadic interaction as shown in Figure 2.3, we gain many opportunities to scaffold the child’s learning process. It is this social role of the robot that we are interested in. Like many others, we believe this may make robots a very powerful tool for learning.

However, we do not yet know how robots may leverage their unique social role to have a meaningful impact—we need to better understand in what ways they may excel and add value to the learning process. Furthermore, in lab settings we can not easily simulate the richness and complexity of social learning as it develops and evolves over time—we need to conduct more longitudinal in the wild studies. For this we need to gain a better understanding of what it takes to develop and use robots in real classrooms for longer periods of time.

The next chapter further discusses the aim and approach we take to address these problems.
The overall aim in this thesis is to gain deeper and richer insights into how children work with social robots as they become embedded in their regular educational activities. As discussed in the previous chapters, robots in educational applications may affect all kinds of aspects related to the learning process and the learner’s enjoyment, engagement, and motivation. For this thesis we look at learning situations where the child learns somewhat independently from their teacher and peers, but can still benefit from some form of social support and scaffolding. Inquiry learning offers a suitable context for investigating this because it can be done with minimal supervision and it offers repeatable, structured tasks of varying difficulty, for which the scaffolding mechanisms are well-defined.

We design and implement a robot-extended learning task and evaluate how children interact with it. We do this from the conviction that the robot is a complete “social package”; it appears anthropomorphic in the way it looks and moves, it expresses meaningful verbal and non-verbal behaviours at appropriate moments, and it shows coherent responses to (social) actions from the user. It is this holistic composition of social capabilities that differentiates the robot from other learning tools. Given this point of view, we investigate how we can leverage such a robot in areas where non-robot learning tools may struggle. However, we need to better understand how to develop robots for this context and how to measure children’s responses to them.

Working towards this, we formulate three main objectives. Figure 3.1 presents the roadmap that outlines the milestones reported in this thesis.

The first objective is to understand the interaction context for the specific task and user group, and construct a platform that lends itself for doing research therein. The robot should navigate the social aspects of supporting and scaffolding the inquiry learning process of primary school children. It should cope with various contingencies that may arise and it should be usable for children working independently. Through an ethnographic study with pairs of children we gain a better understanding of the relevant social interactions of learners in the task (see Chapter 4). In doing so we identify concrete opportunities where the robot may offer a meaningful contribution and we compile guidelines on how to design the robot’s social behaviours accordingly.
A research platform is constructed that consists of appropriate learning tasks and robot behaviours, robust interaction paradigms, and the novel instruments required for measuring how children work with them (see Chapter 5). The platform is geared towards autonomous unsupervised interactions in real classrooms.

The second objective is to show that the platform can be used to design and conduct pertinent studies in the envisioned context and task. With the design guidelines and research platform in place, we should have the tools necessary to run studies that yield relevant results. Through a controlled experiment with brief single-session interactions we show just that (see Chapter 6). Using a mixture of qualitative and quantitative methods we investigate the impact that the robot has on children’s explanations as compared to a non-robot system.

The final objective is to explore how this all works out in the long term, both regarding study effects of objective 2 and the suitability of the research platform of objective 1. By placing the robot and educational materials in real classrooms we stand to gain a better understanding of the effects that it may have on the children. We are particularly interested whether beneficial effects of the robot are sustained across repeated interactions, and what happens with the interactions as children become accustomed to the system. Data collected during a longer-term deployment in the wild is analysed from various viewpoints. We investigate how children’s mindset towards learning is shaped by the robot (see Chapters 7), and explore how children’s learning process develops over time (see Chapter 8).
“Good morning kids!” “Good morning Mr Donovan!” the kids respond cheerfully. “What is that?” a boy asks, while pointing to the large box on the teacher’s desk.

“Ah, I have some exciting news for you! Starting today we have a new classmate.” Donovan smiles, “And not just any classmate, this is a special robot that can help us when we’re stuck on a task.” Donovan unpacks CHK-e from the box, places him on his desk and presses the power button.

“WOOOAH!” They gasp in amazement as CHK-e opens his eyes and takes in his new home. Children gather around to take a closer look.

“He is so cute!” someone says, while touching the robot’s hand. “What is his name?” someone else asks. They spell out the model number printed on his chest, “C-H-K-e... CHUCKY! His name is Chucky!”

“Hahaha well I guess it is!” Mr Donovan replies. “Now kids, I first need to read this manual to get started. I’ll leave Chucky here for now, so he can get used to his new environment. You lot can start with your first assignments for the day.”

“Chucky...” That has a nice ring to it, Chucky thinks to himself. “A friendly name, totally not scary at all.” He watches with fascination as the children get to work.

Two kids at the front of the class, Mees and Sophie, take a large box labelled ‘Buoyancy’ from the cupboard. Mees fills a tub with water while Sophie starts lining up various boats, rafts, and ballast objects. “Which one do you think will float the best?” Mees asks. “Not sure, but I think that one will sink the hardest!” Sophie replies, as they start working their way down the listed steps of the assignment.
4.1 Introduction

One of the objectives of this work is to develop robots that can support children’s learning process. To achieve this, we must design robot behaviour and human-robot interaction patterns that are appropriate for the learner and the educational context. Generalised pedagogical and educational psychology theories inform us about what we might generally expect by way of interactions that occur between learners. However, they do not specify exactly when and how these interactions materialise in practice, nor do they specify in what way such social interactions carry across to child-robot learning situations. To ground the design of actual behaviour and interaction of the robot we also need real examples of behaviours exhibited by children during situations in which the robot should operate. From such examples we can derive concrete opportunities and guidelines for when and in what form exactly robots may contribute to a child’s learning.

In this chapter we report on a contextual analysis that was carried out to derive guidelines for the design of social robots to support primary school children in a collaborative inquiry learning assignment. This work was guided by the following research questions:

RQ1 – Which typical behaviours and social interactions do we identify in learners working with our assignment and how do these behaviours align with theory?

RQ2 – How do children progress through the various phases of the learning assignment and do the types of learning behaviour they exhibit vary with these phases?

RQ3 – What are opportunities for social robot interventions in regard to interaction patterns, specific behaviours and roles, to contribute to the child’s learning in the assignment?

Section 4.2 describes the approach of the contextual analysis, in which we annotated video recordings of pairs of children collaborating on a learning assignment in their school. Section 4.3 presents the annotation schemes we used to analyse these video recordings, the results of which are presented in Section 4.4. Based on these results, we formulated four generic recommendations and fourteen concrete design guidelines for robot behaviour, presented in Section 4.5. As part of a subsequent chapter we discuss how these guidelines were used to construct four variants of child-robot learning scenarios. There, we also reflect anecdotally on the social learning behaviour exhibited by children interacting with these variants.

4.2 Contextual analysis

We envision that a social robot is able to navigate the social space of the educational, collaborative, and relational levels of the ECR model by taking on a variety of roles in the collaborative learning process. In Chapter 2 we explored examples of related work, which helped us to highlight and illustrate possible contributions of educational robots. In summary, literature describes various (social) roles that an educational robot could have: explaining to the learner, be explained to by the learner, encouraging the learner, praising the learner in various ways, giving good or bad examples as a fellow learner, and many more. The effectiveness of these actions depends to a certain extent on the relation between the learner and the robot, and how the robot behaves and presents itself to the learner. As input to the design guidelines we carried out a contextual analysis to determine concrete opportunities for when and how robots may contribute to a child’s learning process.
4.2.1 Method

The field of Human-Computer Interaction often emphasises the importance of considering the context in which interactive products are to be used (e.g. [16, 17, 103, 140]). Over the years different terms have been used for contextual analysis: “Context-of-use analysis”, “Usability Context Analysis”, or “Contextual Design”. Assessing and accounting for contextual factors is considered as an integral part of a user-centred design process aimed at optimising usability of products. By mapping the challenges and opportunities that exist in a certain context, the outcomes of a contextual analysis will typically have a drastic influence on the design of an interactive product. Therefore, a contextual analysis should take place at an early stage—before having a fully developed prototype or product.

The method of contextual analysis aims to take into account, as much as possible, all factors that may influence the use of a product. Factors include: characteristics of the user group; tasks that users aim to carry out with the product; and the technical, organizational, and physical environment where the product is to be used. A contextual analysis typically consists of approaches like stakeholder analysis, surveys, and focus groups [140], as well as ethnographic strategies like case studies and field studies [146] (or combinations thereof).

In the contextual analysis we used such ethnographic methods to investigate the context in which children learn with our specific inquiry assignment. We conducted a field study where we recorded pairs of children working with the assignment. We then used a grounded theory approach to design an annotation scheme with which we analysed the observed behaviours [149, 180].

4.2.2 Venue and participants

Since we are interested in gaining insights in typical behaviours of children as they occur in practice, we aimed to maintain a high degree of ecological validity throughout the study. Therefore, we focused on pairs of children working on a learning assignment in their own school during regular school hours, without robot. A total of 22 Dutch Montessori elementary school students (50% female), aged 6-9 years old (the target age range of the EASEL project) participated in the study. Children in this age range show a variance of cognitive development, verbal, and non-verbal behaviour. As such, including children across this age range will give us a broad spectrum of insights on possible behaviours. The school is located in a residential area near the University of Twente, in the city of Enschede, The Netherlands.

4.2.3 Procedure

Prior to the study, ethical approval was granted by the ethical boards of the Behavioural, Management and Social sciences (BMS) and the Electrical Engineering, Mathematics and Computer Science (EEMCS) faculties of the University of Twente. Permission forms were distributed to the children’s parents through the school’s regular communication channels.

The experiment started with an explanation to a pair of children about the assignment. The researcher showed them the learning materials and demonstrated how they worked. This introduction took approximately 5 minutes. After the introduction, the children would each receive the assignment sheets with five tasks from the
Chapter 4

expirmenter. The children were given approximately 30 minutes to complete the assignment. When a pair finished the assignment, the experimenter would thank them for their participation and say goodbye. During the experiment, the experimenter was present so that children could ask for help if they had any questions.

The interactions were recorded by three cameras: two for recording actions of each of the two children and an overview camera to record the learning materials.

4.2.4 Balance scale materials

The physical balance scale instrument was originally designed by Inhelder and Piaget [82]. In our study, children used such a balance scale (see Figure 4.1) to explore the moment of force. The moment of force (or torque) is a physical concept that describes the rotational force acting upon an object in relation to a reference point. To explore this concept, children did small experiments in which they placed pots of various weights at various distances from the scale’s central pivot point.

The two-armed wooden balance had a central pivot point and three pins at equal increasing distances on each arm. These pins were numbered 1 to 6 from left to right. Numbers 1 to 3 were given a blue label and numbers 4 to 6 were given a green label. The children received four pots that could be placed over the pins: two heavy red pots and two lighter yellow pots. Two wooden blocks could be placed under the balance to secure it in a horizontal position.

4.2.5 Learning assignment

Children received a total of five numbered tasks as part of their learning assignment. Task instructions were printed on paper and consisted of brief textual descriptions and illustrations.

Task 1, 2, 3, and 4 were similar in nature. In these tasks the children’s instructions and actions were symmetrical and synchronous: each child had the same task instructions and could complete the same actions. Each task was structured according to the steps of the inquiry learning. In the preparation step, children were first asked to place a wooden block underneath each arm to prevent the balance from tilting prematurely. They were then asked to place two pots on the balance (pot colour and position were given explicitly in each task). In the prediction step, children were asked to predict what they thought would happen when the wooden blocks were to be removed. They could select one of three options: 1) the blue side will go down; 2) the balance will stay in equilibrium; or 3) the green side will go down. In the experiment step, children
were asked to remove the wooden blocks, allowing the balance to tilt freely. During the observation step they explicitly stated their observation about the position of the balance and explained whether this was in agreement or disagreement with their initial prediction. Finally, in the conclusion step they would give an explanation for their observation using their own words (explain why the balance tipped left, right, or stayed in equilibrium). They gave this explanation verbally to each other and the experimenter.

Task 5 was structured differently in order to possibly trigger more diverse collaborative behaviours. In this final task children worked asynchronously and were given asymmetrical task instructions: here it was necessary for the children to take consecutive steps, where each child took different actions. First, child A would choose one of the pots and place it on the blue side of the balance. Then, child B would choose one of the pots and give it to child A. Subsequently, they had to answer the following question together: “where should the pot be placed on the green side to keep the balance in equilibrium?” Child A would then place the pot on the green side where they thought it should be placed to keep the balance in equilibrium. Finally, after removing the blocks, the children would explain to each other if their prediction was correct and why the balance was in this position.

4.3 Annotation scheme

To further understand the children’s educational, collaborative, and relational interactions and to understand how the children progress through the phases of the inquiry cycle, we manually annotated the video recordings using the ELAN tool\footnote{The ELAN tool (https://tla.mpi.nl/tools/tla-tools/elan/) was developed by the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands} [194]. The resulting label sets in the annotation scheme are presented in sections 4.3.1–5.

The annotation scheme for the labels under the educational, collaborative, and relational levels of the ECR model was developed through an inductive grounded theory approach, as illustrated in Figure 4.2. Salient events were extracted from sample fragments of the recordings through open coding, to get a sense of the kind of behaviours that could be found in them. Through iterative axial coding these salient events were renamed and grouped into emerging coherent themes, which were used to derive explicit label sets. Events that could not be cleanly grouped in a theme, but were nonetheless deemed important to the learning process, were labelled as Supplementary interactions. Finally, the levels of the ECR model were used as a lens through which to align all labelled interactions. No clear labelled themes emerged from the data on the relational level, due to the highly diverse nature of the relational interactions that took place. However, the category supplementary interactions was found to contain distinctive events and interactions aligning with the relational level.

The labels for the five inquiry processes were deduced from literature. Starting with an open coding approach we extracted concrete events and interactions from the recordings. These were then labelled according to the corresponding inquiry process. Using the developed annotation schemes, coder A labelled 100% of the data. To calculate the inter-annotator agreement score, coder B labelled an overlapping 20% of the data. On the combined annotations of the educational and collaborative levels, we computed a Cohen’s Kappa of 0.11. Inspection of the disagreement data showed that this low score was mostly caused by one coder assigning a label to a certain video fragment while the other coder did not. Upon discussion of these fragments, it became
clear that the majority of the behaviours that were missed by one of the coders, did indeed rightfully belong under the annotated label that the other coder assigned to it. Similarly, when we excluded these annotation instances that were missed by one of the coders, Cohen's Kappa increased to 0.77. The Cohen's Kappa for all annotations in the inquiry process category was 0.61. After excluding the fragments that were missed by one of the coders, Cohen's Kappa increased to 0.91. These inter-annotator agreement scores left us confident that the labels that we gathered from the annotation procedure were meaningful enough for further analysis, even though a number of instances of relevant interactions might not have been assigned a label.

The results of these annotations will allow us to identify how children express themselves when navigating various roles in relation to their learning partner. Such insights will form the basis for behaviour design guidelines for a social robot that will function in similar roles. Additionally, these annotations will allow us to identify when children express certain behaviours, which gives us clues to when the robot might contribute. Finally, we will be able to assess whether our specific learning task is appropriate for the target user group, and whether it lends itself to investigate the full inquiry learning cycle.

The annotation schemes are described in more detail in the next five sections.
4.3.1 Label set for educational interactions

During the course of an assignment, children often engaged in interactions that were educational by nature. That is, they engaged in discussion and knowledge construction around the learning content. On the one hand, we observed very explicit educational exchanges (for instance, child A asking child B for an explanation or opinion). On the other hand, children engaged in implicit educational interactions (for example, child B integrating an argument in their explanation that was previously made by child A). Additionally, we observed instances where children would engage in discussions where they agreed or disagreed with each other’s answers or actions.

The following labels were used to identify the interactions on the educational level. These labels were partly based on work by Van Dijk et al. [55], who investigated collaborative drawing behaviours.

Elicitation/Asking questions: When children want to extract knowledge or support from their partner, mostly in the form of a question. For example, “What do you think will happen?”.

Explaining: When a child gives any form of verbal explanation directed at the partner about the assignment or the principles of balance. This explanation may be spontaneous or elicited. For example, “This pot is at the same spot as the other, so I think it will stay in equilibrium”.

Thinking aloud: This label includes cases where a child thinks aloud about the construction of their argumentation, but this externalisation is not explicitly directed towards the partner. Such utterances are often less elaborate than explanations. For example, “hmm... I think... hmm... the left side will go down”.

Agreement or disagreement: Instances where a child shows basic (dis)agreement with their partner’s reasoning, without explicitly showing comprehension. This includes, for example, verbal utterances (“yes, I think so too” or “no, that’s not right”) and non-verbal gestures (nodding or shaking their head).

Integrating: This label goes beyond basic agreement, when a child uses the content of a partners’ explanation in their own line of argumentation. For example, child A says: “I think the balance tipped left because the pot on the blue side is placed more towards the end” and child B responds: “Because it is more towards the end of the balance the left side is heavier”. In this case, the explanation of child A is labelled as Explaining and the response of child B as Integrating.

Elaborate critique: When a child criticises or corrects the content of the explanation of the partner, going beyond basic disagreement. For example, child A explains “The yellow pot is heavier”, and child B responds “No, it’s lighter!” the response of child B is annotated as Elaborate critique.

4.3.2 Label set for collaborational interactions

While working on their assignment, children had to collaborate in order to share the physical learning materials. Doing so effectively required them to align their interpretations of task objectives, communicate their approach, and agree upon actions. The following labels were used to identify the interactions on the collaborational level.
Reading aloud: When a child reads a task or instruction out loud. Since it is often difficult to distinguish between reading out loud for oneself and reading for the partner, this label includes both cases.

Coordination: When a child makes explicit suggestions or explanations about how to proceed with the task, or explicitly mentions their progress within the task. For example, “it’s your turn to place a pot” or “I’m finished with this one”.

4.3.3 Label set for relational interactions

In order to collaborate effectively, partners needed to maintain mutually benevolent attitudes with respect to each other. Their attitudes were shaped and maintained through relational interactions, which took the form of shared experiences, emotions, and frustrations. Due to the diverse and distinctive nature of these relational interactions there were no clear labels that emerged from the data through the grounded theory approach. Instead, relevant events from the Supplementary interactions category were used to generate a first exploratory overview of the relational interactions.

4.3.4 Supplementary interactions

We observed many salient behaviours and interactions that appeared to be relevant to the above label sets, but did not fit cleanly into any of the available labels. These spanned a wide range of actions. For example, at one point a child grabbed and took materials from another child; this action happened only in one session, but was considered very salient for collaborational interaction. Therefore, to fully capture the richness of the social interactions we labelled such additional observed behaviours as Supplementary interactions. Behaviours with this label were then ordered along the three levels of the ECR model and are reported and discussed as such in the respective results sections below.

4.3.5 Label set for inquiry processes

A typical inquiry learning cycle is composed of five processes: 1) preparation; 2) hypothesis generation; 3) experimentation; 4) observation; and 5) conclusion. To analyse how children progressed through these phases, we identified behaviours that represent these processes using the following labels:

Preparation: When children prepare the physical learning materials in accordance to the task, e.g. adding pots on the balance and placing the wooden blocks.

Hypothesis generation: In this phase, children state their hypothesis, e.g. by verbalising out loud to each other, or by selecting a hypothesis on their assignment sheet.

Experimentation: Performing the actual experiment, i.e. removing the wooden blocks so that the balance tips over.

Observation: In this phase children state their observations about the experiment, by verbalising or pointing (e.g. “That side went down!”), or by documenting the observation on their assignment sheet.
Table 4.1: Frequency of educational interaction labels.

<table>
<thead>
<tr>
<th>Educational label</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elicitation/Asking questions</td>
<td>116</td>
<td>25.3%</td>
</tr>
<tr>
<td>Explaining</td>
<td>164</td>
<td>35.8%</td>
</tr>
<tr>
<td>Thinking aloud</td>
<td>53</td>
<td>11.6%</td>
</tr>
<tr>
<td>(Dis)agreement</td>
<td>97</td>
<td>21.2%</td>
</tr>
<tr>
<td>Integrating</td>
<td>6</td>
<td>1.3%</td>
</tr>
<tr>
<td>Elaborate critique</td>
<td>22</td>
<td>4.8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>458</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

**Conclusion:** When children draw conclusions about the current task, by explaining to each other. This phase only occurred verbally, e.g. by explaining whether their hypothesis was correct and why.

### 4.4 Results of the contextual analysis

This section reports the content of the annotations. All the 11 pairs of children completed tasks 1, 2, 3, and 4, each of which followed a similar structure. However, in 36% of the tasks they needed some assistance from the researcher. In most cases, this simply meant the researcher reading the task out loud or rephrasing the task instructions in different words. In cases where that did not help, the researcher would ask thought-provoking open questions without revealing the answer.

Task 5 took on a different structure, which likely increased its complexity. During this final task, children required more support from the researcher; only 3 out of 11 pairs of children were able to complete the last task without any help or guidance. The children expressed confusion about the structure of this task, which was composed of different steps than the previous tasks. In particular, they often did not understand the different responsibilities that they needed to fulfil in order to complete the task. Still, during this task we did observe interactions related to learning.

#### 4.4.1 Educational interactions

In identifying the educational interactions between children, a total of 458 annotations were performed. Frequencies for individual labels on the educational level are shown in Table 4.1.

The most prevalent labels were *Elicitation*, *Explaining*, and *(Dis)agreement*, which indicates that children were intentionally helping each other by explicitly sharing knowledge. Their helping behaviours focused primarily on topics related to the process (e.g. “I don’t understand what we have to do”) and the learning content (e.g. “I don’t understand what happened”). Additionally, learners unintentionally helped their partner during behaviours labelled as *Thinking aloud*, by implicitly sharing knowledge about the task or content. Finally, we sporadically found instances where children explicitly integrated or critiqued their partner’s explanation or reasoning process, illustrating a deeper level of discussion and debate.

In addition to the annotations labelled according to the annotation scheme, we performed an explorative analysis of the annotations that were labelled as *Supplementary interactions*. Firstly, in these annotations we observed many instances of
Table 4.2: Frequency of collaborational interaction labels.

<table>
<thead>
<tr>
<th>Collaborational label</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading aloud</td>
<td>23</td>
<td>30.3%</td>
</tr>
<tr>
<td>Coordination</td>
<td>53</td>
<td>69.7%</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>100%</td>
</tr>
</tbody>
</table>

From these additional observations we found that children discussed their observations and updated their knowledge while progressing through tasks. We also found that gestures and gaze behaviours played an important supplementary role in the collaborative inquiry learning interactions: children seemed to naturally take advantage of the physical nature of the embodied learning materials when explaining to their partner.

4.4.2 Collaborational interactions

In identifying collaborational interactions between children, a total of 76 annotations were performed. Frequencies for individual collaborational labels are shown in Table 4.2. In general, we found behaviours corresponding to all of the criteria, requirements, and choreography of joint activities discussed in Section 2.4.2.

Behaviours that fell under the label Reading aloud included children reading (parts of) the assignment out loud. Often, this role was taken on by the child who was most proficient in reading. However, when both children had difficulties reading the task instructions, the experimenter would read the assignment out loud. By reading instructions out loud, the children seemed to ensure that both team members were in agreement on the task objectives. Additionally, this naturally provided whoever was pointing and gesturing behaviour. This was often used as a non-verbal clarification or explanation (e.g. pointing at a pot or location on the balance, or gesturing the position of the balance, or pointing at other’s assignment sheet to indicate the correct answer). Secondly, we found several implicit and explicit indications that children were comparing and aligning their own answers to those of their partner. For instance, they would look at and sometimes copy the answer of their partner, they would discuss explicitly (e.g. “What are you going to fill in?”), or they would announce when they had a breakthrough of their own (e.g. “I know the answer!”). Thirdly, although most children accepted lessons learnt through observations of the embodied learning materials (e.g. “I have now of course seen it... so I have to believe it”), we occasionally saw instances of confirmation bias. For instance, some children would alter reality to fit their prediction (e.g. perhaps jokingly, they would say “See, I was correct, I said it would stay balanced”, while artificially holding the balance in equilibrium with one hand). Others would adapt their original (incorrect) hypothesis to match their observations on the fly (e.g. “But I did think this would happen beforehand” or “You know, actually I knew this was going to happen”). Nevertheless, in some instances this prompted children to discuss an alternative hypothesis or propose an alternative experiment (e.g. “What would happen if we would place this pot here and the other here?”). Finally, some children showed curiosity about the balance unrelated to the assignment, which provoked them to investigate beyond the scope of the subject matter (e.g. they would try to figure out what was inside the weighted pots or they would place unrelated items on the balance).

From these additional observations we found that children discussed their observations and updated their knowledge while progressing through tasks. We also found that gestures and gaze behaviours played an important supplementary role in the collaborative inquiry learning interactions: children seemed to naturally take advantage of the physical nature of the embodied learning materials when explaining to their partner.
listening, the opportunity to ask clarifying questions, or suggest a plan of action.

Among behaviours labelled as Coordination, we mostly found interactions that were related to the planning and execution of individual actions. For instance, children would take turns manipulating the learning materials (e.g. “Now it’s your turn to place the pots”), or they would share responsibility (e.g. “I’ll place this pot, you do that one”). Furthermore, they would occasionally plan ahead based on the structure of the task (e.g. “Hang on, before we can remove the blocks we first have to give our answer”). Finally, we identified instances where children explicitly mentioned their progress in the assignment (e.g. “There, that was task 2 done, let’s move on to the next!”).

From the explorative analysis of the behaviours in the supplementary interactions category, we identified several additional situations that were relevant on the collaborative level. Firstly, in addition to the explicit verbal interactions coded under the Coordination label, we found several instances of successful implicit coordination. While children worked together on the physical learning materials, they would often maintain a shared focus of attention through mutual gaze and by observing each other’s actions. Additionally, we noticed that children would often use behaviours, body language, pointing gestures, or gaze to signal their intentions to their partner. For example, children would seek for approval or permission while doing actions (e.g. very slowly/ hesitantly starting an action, while searching for eye contact, giving their partner opportunity to intervene), or they would assist their partner’s actions (e.g. holding the balance steady while the partner removes or places pots). Secondly, we observed instances where misalignment of actions or goals occurred. Children could often explicitly signal such a situation to their partner verbally (e.g. by stating “This pot is in the way” or “I think we did it wrong”) or non-verbally (e.g. obstructing or overruling their partner’s actions, for instance by forcefully taking or removing a pot). In cases like these, partners would seek re-alignment through more explicit coordination. They would, for instance, seek agreement with their partner before performing actions (e.g. “Can I do this now?” or “Are you choosing that pot..? Ok good!”) or they would explicitly time and synchronise their actions (e.g. “Now? Yes? ... Yes now!” or “3..2..1..Now!”). Finally, there were instances where misalignment of task progress occurred. For example, children would often check their partner’s progress in the task by looking at and comparing each other’s assignment sheets. If they noticed that their partner was falling behind, children would often use either of the following two strategies: 1) they would either dawdle patiently while waiting for their partner to catch up, not directly contributing to their partner’s progress (although some children asked the experimenter “Do we need to wait for each other?”). Or they would 2) directly interfere with their partner’s progress by prompting the answer (e.g. “Fill this one in [pointing at answer sheet]”), or by taking on a more dominant role where they would give their partner explicit tasks or would explicitly direct the actions of their partner (e.g. “Now you do this” or “You have to do that”).

4.4.3 Relational interactions

Since we did not have a grounded annotation scheme for behaviours that constitute relational interactions, we looked at the interactions that fell under the supplementary interactions category. In doing this, we identified several situations that appeared relevant on the relational level. Firstly, in addition to mutual eye contact and shared focus of attention, children would often share mutual emotions while progressing through the tasks. For example, they would smile and laugh together when something
Table 4.3: Frequency of inquiry process labels.

<table>
<thead>
<tr>
<th>Inquiry process</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>316</td>
<td>37.4%</td>
</tr>
<tr>
<td>Hypothesis generation</td>
<td>94</td>
<td>11.1%</td>
</tr>
<tr>
<td>Experimentation</td>
<td>196</td>
<td>23.2%</td>
</tr>
<tr>
<td>Observation</td>
<td>150</td>
<td>17.7%</td>
</tr>
<tr>
<td>Conclusion</td>
<td>90</td>
<td>10.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>846</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

(unexpected) happened, they would celebrate or cheer together when they had made a correct prediction (e.g. “Yaaay!”), or they would explicitly share their confusion or frustration when their predictions were proven wrong (e.g. “...HUH...??!!”). In addition, some partners would occasionally mimic actions of the other for comic effect, often resulting in shared laughter (e.g. task related actions such as placing pots in mirroring locations, or non-task related actions such as stacking papers). Secondly, on the one hand we observed examples of friendly courtesy towards each other (e.g. “[handing over a pot] if you want, you can place this one” or “you may remove the blocks if you like”). On the other hand, we saw signs of frustration about their partner’s actions, often expressed as deep sighs or rolling their eyes. Finally, we observed cases where partners engaged in mutual “mischief” outside of the scope of the task, where they would reinforce each other’s mischievous behaviours.

The behaviours that we identified on the relational level suggest that children worked towards establishing and maintaining a positive relationship. Mutual shared experiences (both positive and negative) contribute to a stronger social bond, which in turn can strengthen the feeling of being “allies in learning”. Although we did observe instances of frustration, overall the children acted mutually benevolent towards each other. Additionally, we anecdotally identified behaviours that could be associated with building rapport. For example, children mirrored or mimicked each other’s behaviours, gestures, and facial expressions, suggesting a level of interpersonal synchrony.

4.4.4 Inquiry processes

In identifying the inquiry processes children engaged in, a total of 846 annotations were performed. Frequencies for individual labels are shown in Table 4.3.

These results verify that children go through all phases of the inquiry learning process when working with the balance task. However, we found an imbalance between the frequency of observations in the various phases: children more frequently engaged in preparation, experimentation, and observation behaviours. In fact, within a single task, children would often complete multiple quick, successive cycles of preparing and experimenting, skipping the hypothesis generation and conclusion phases. Preparation, experimentation, and observation phases were composed primarily of actions and verbalisations related to the practical or physical nature of the task (e.g. placing/removing pots and blocks, or remarking on the tilt of the balance). On the contrary, behaviours that were associated with the hypothesis generation and conclusion phases were more cognitive and reflective in nature and focus on constructing and organising knowledge (e.g. discussing or explaining about how to interpret the experiment outcome, or combining prior knowledge and new knowledge to construct an updated hypothesis).
4.4.5 Conclusion of the contextual analysis

The analysis revealed a rich collection of behaviours aligned with the ECR model that children exhibit in practice, in the context of our specific learning assignment. Furthermore, we found that children went through all phases of inquiry and displayed learning behaviours appropriate to each phase. These results show firstly, that this inquiry learning assignment is suitable for conducting research on collaborative interactions. Secondly, they show that the assignment is usable for the children in the target user group.

Based on these findings from the contextual analysis we identified specific opportunities for when and how a social robot may make a meaningful contribution to the learning process. In the following section we present concrete design guidelines that focus on specifying potential roles and behaviours for the robot, as well as opportunities in the learning cycle where a robot may potentially intervene.

4.5 Design guidelines for educational robots

In this section we present four general recommendations and fourteen concrete Design Guidelines (DG 1-14) for developing social robot behaviours, derived from the findings in the literature and the contextual analysis. Furthermore, based on observations from child-child interactions we speculate about examples of possible robot behaviours and phrases that illustrate how each design guideline may materialise in practice.

**Recommendation 1 - Designing educational behaviours: The robot should offer appropriate content support and process support**

Learners who construct knowledge [167] through the process of inquiry learning [105] require a certain degree of scaffolding to achieve their next level of cognitive development [213]. In the observations from the contextual analysis we found many instances where children scaffolded each other’s learning process. In some instances, we see behaviours that don’t contribute much to learning, such as directly prompting answers or even blatantly copying answers. However, in most cases scaffolding takes place through explanations from one child to the other.

**DG 1: The robot could use the physical learning materials to supplement its scaffolding**

From the observations of children interacting we found that, when explaining, they naturally used the physical learning materials as tools or props to supplement their explanations. A robot could take advantage of similar techniques when scaffolding the child’s learning.

Both the robot and the learning materials are physically embodied in the environment of the child, which can be quite naturally leveraged to create more coherent scaffolding behaviour. By referring to objects or aspects of the learning materials, the robot can further supplement its explanations or increase the legibility of its instructions. Such supplementary didactic behaviours may take several forms depending on the robot’s embodiment and physical capabilities.

For example, the robot may use non-verbal behaviours like gazing, gesturing, or pointing to reference locations, objects, or components of the learning materials. These behaviours can be used while the robot is explaining or reading out loud, or when
objects are referenced by the child or assignment. Furthermore, robots with fine motor skills may physically manipulate learning materials, to offer an illustration (e.g. the robot may pick up a red pot and a yellow pot and say “this one seems heavier, what do you think?”).

**DG 2: A robot should explicitly take on a certain role**

Based on the observations we found that an individual child might be more knowledgeable in one aspect of the task while they are less knowledgeable in another. Furthermore, we saw that children naturally and effortlessly switched between roles as they progressed through the learning tasks together. In the current analysis we have not investigated in detail how children establish and regulate their respective roles of more knowledgeable and less knowledgeable team member—a process which seems to occur implicitly. However, we do see opportunities for a robot to support the learning process by explicitly navigating a certain role.

For a robot to effectively help children learn, the child must know how they can benefit from the robot’s presence. In other words, the types and forms of support (e.g. actions, information, knowledge) that the robot will be able to offer them will depend on the role that the robot takes. Since we are unsure how the robot may navigate between roles implicitly, we suggest that the robot should have explicit means of presenting itself in a certain role, in order for children to more easily appraise the support that they may potentially receive from the robot.

For example, a robot could explicitly introduce itself in a certain way or the designer could provide the robot with a certain background story. On the one hand, a robot navigating the role of a less- or differently knowledgeable other, could for instance, introduce itself: e.g. “Hi, my name is […]. I’m interested in learning about topic X, would you like to investigate it together?” or could be introduced by the designer with a background story: e.g. “This is […], an exchange student from the robot school. Like you, it is studying topic X, maybe you can study together.” On the other hand, a robot navigating the role of a more knowledgeable other, could for instance, introduce itself: e.g. “I have already learned something similar to this topic, maybe I can help you with it.” or the designer could describe the background story of the robot: e.g. “The robot has practised this task a lot of times before, so you can ask it for instructions.” The choice for which role fits best for the robot will likely depend upon how the designer wishes to present and introduce the robot to the child, as well as the looks of the robot, the type of task, the learning context, and the research interests.

**DG 3: A robot operating in the role of differently knowledgeable or less knowledgeable other could leverage the interactive explanation paradigm**

In the observations we often saw educational interactions by the more knowledgeable child being triggered by something that the less knowledgeable child said or did (e.g. making mistakes, looking confused, or asking). The more knowledgeable child would then use their expertise to provide support to the other child (e.g. correcting, explaining, or thinking/reading aloud).

In such situations a peer-like robot can offer social facilitation (i.e. audience effect) through which it can leverage the interactive-explanation or the learning-by-teaching paradigm [168] by prompting task-related questions to provoke a (verbal) response.
This may trigger the child to (re)organise specific elements of their knowledge to offer a coherent explanation.

For example, in some cases, the robot may ask questions that require the child to reflect on the process of the learning assignment (e.g. “What did we do in the last step?” or “What does the task say about the next step?” or “What was our initial hypothesis, was it correct or not?” or more generically, “What happened to the balance?”). In other cases, the robot may prompt for specific knowledge that children should have previously acquired, or highlight information that they need to answer the current task (e.g. “Can you remember which of the pots was heaviest?” or “Why is that pot further away from the centre than the other pot?” or “In the last exercise the balance tipped to the left, so why does it stay balanced now?” or “What do you think would happen if we moved one of the pots closer to the pivot?” or more generically, “Can you explain why the balance behaves like this?”).

**DG 4: A robot operating in the role of a more knowledgeable other could offer tutoring to the learner**

Observations revealed many instances where children needed process and content support to understand the learning assignment. In many cases such support was offered by the more knowledgeable child. The experimenter would offer additional tutoring when both children were unable to continue with the task.

In these cases, a tutor robot may offer the child directions, advice, and explanations regarding the learning content and learning process. It may help them in specific areas where they are struggling or where they require additional information. Depending on the robot’s reasoning capabilities it could either attempt to automatically recognise when a child requires tutoring or could respond to a child’s explicit request for help. In some cases it would be sufficient for the robot to read the instructions out loud and in other cases the robot may need to employ more advanced forms of tutoring.

For example, regarding the inquiry learning process, the robot may offer additional background information about the steps of inquiry (e.g. “In the hypothesis step we must think about what we expect will happen” or “When giving a conclusion, try to think about whether your hypothesis was correct or not” or simply reading the assignment out loud). Regarding learning content the robot may offer, for example, analogies that the child might be familiar with (e.g. “This balance is like a see-saw on the playground. What happens if a mouse and an elephant sit on either side of a see-saw?”), or it might refer to prior knowledge (e.g. “Remember, we saw last time that the red pot is heavier than yellow”), or might explain underlying conceptual principles (e.g. “When a pot is further away from the pivot it puts more force on that side of the balance”).

**Recommendation 2 - Designing collaborational behaviours: The robot and child should engage in joint activities**

Effective collaborational interactions require collaborators to work towards aligning their goals, objectives, strategies, abilities, and actions [40, 107]. According to Klein and Feltovich [107] successful joint activities are achieved (among other things) through an interplay between the following five aspects: a) an intention to collaborate; b) a level of interdependence between collaborators; c) common ground; d) directability of one’s actions by their partner; and e) effective signalling behaviour. In observations we saw each of these aspects materialised in one form or another in
the children’s behaviours and interactions. A social robot aimed at collaborating with a child should be designed to also navigate these aspects of joint activities. In the next five design guidelines we discuss each of Klein and Feltovich’s requirements and translate them into more specific robot behaviours.

DG 5: To express an intention for collaboration, the robot should disclose its own goals and should discuss task objectives, strategies, and joint actions

Klein and Feltovich [107] describe the intention to collaborate as a mostly implicit and unspoken agreement between all team members. From the observations, however, we found examples where children agreed upon goals and strategies, and examples where children discussed intentions to perform individual actions. By doing so they explicitly expressed their intention to collaborate, in addition to any implicit unspoken agreement that might have preexisted.

Similarly, we suggest that a robot should express such intentions for collaborating. Since we did not investigate unspoken intentions, we can only offer suggestions for explicit robot behaviours. It is important to note that educational robots are usually created and programmed to accomplish a certain objective (e.g. to improve the child’s learning in some form or another). This objective is not necessarily known to the child. However, to signal to the child that the robot has a benevolent intention to collaborate, it could disclose a different “personal goal” (e.g. with an introduction or a background story). While working on the learning assignment with the child, the robot should continue to express its intention for collaboration by discussing task objectives, strategies, and actions. The robot may take various approaches, depending on the type and content of the assignment, the learning context, and the role of the robot in relation to the child.

For example, a more knowledgeable or tutoring robot may introduce its personal goal to the child as: “teaching subject X”. During the rest of the interaction, that robot could then act as a leader, tutor, or expert: reading the instructions out loud to the child (e.g. “The task says we should do: X..Y..and..Z”), referring explicitly to the task instructions when they are violated (e.g. “I think we’re doing X wrong”), and proclaiming progress when the instructions are followed (e.g. “We finished X, now we do Y to move on to the next step”). However, a less knowledgeable or differently knowledgeable robot may tell the child that its personal goal is: “wanting to learn about subject X”. That robot could then act like a peer-learner during the remainder of the interaction: asking the child to read the assignment out loud, asking explicitly whether the current step meets all task instructions (e.g. “Did we do everything that the task says?”), ask explicitly about the next actions (e.g. “What should we do next?”), and inquire about progression through tasks (e.g. “How far are we? When can we move on to the next step?”).

DG 6: To support a level of interdependence, the robot and child should facilitate each other’s actions

Interdependent activities, according to Klein and Feltovich [107], are interwoven actions between collaborators; their actions mutually influence and affect each other. The learning assignment inherently imposed a certain level of interdependence. Since both children had to share the same physical learning materials, actions from the one, naturally influenced the other. In observations we mostly saw examples of children working together (e.g. assisting each other with placing pots). However, we also
saw instances of children working in separation (e.g. each doing “their own task” in sequence) and even obstructing each other (e.g. taking away pots that their partner had just placed).

A robot that aims to collaborate with children should aim to engage in interdependent interactions. Interdependence in collaborative learning assignments may be introduced in various ways. Firstly, the designer of the learning assignment may choose to impose explicit interdependencies between the child’s and the robot’s actions. Secondly, the robot and child may assist each other’s actions even if this is not imposed explicitly by the structure of the learning assignment.

For example, the assignment might require two different actions to be performed at the same time, or might require a sequence of actions (e.g. first the robot must complete action A, then the child must complete action B, only then can the robot complete action C, and so on). In such situations the robot may either take on a dominant role (e.g. “It’s your turn to do action X”) or a submissive role (e.g. “What should I do next?”). In other cases, the robot may ask the child for assistance when it is incapable of performing a certain action (e.g. asking the child for an object that is outside the robot’s reach). Furthermore, depending on the robot’s physical capabilities it may also attempt to assist the child’s actions (e.g. handing over/pushing objects to the child or manipulating other aspects of the learning materials).

DG 7: The robot should build and maintain common ground with the child in the context of the learning assignment

Klein and Feltovich [107] describe common ground between collaborators as the shared mutual knowledge, beliefs, and assumptions, which continuously evolve during interactions. In general, a strong common ground can result in more efficient communication and collaboration during joint activities. We speculate that, among other things, our children’s implicit mutual gaze, attentive gaze, and shared focus of attention contributed to their common ground. Additionally, we saw instances where children sought explicit alignment by discussing task instructions and actions.

A strong common ground in child-robot interactions may contribute to an increased legibility and predictability of the robot, enabling the child to more accurately anticipate and infer the robot’s future actions. In turn, this increased predictability may enable the child to align their own actions and react to the robot accordingly. Children from the same class might have a level of established common ground, allowing more implicit forms of alignment. However, we expect that not to be the case during initial child-robot interactions: the child may not be aware of the robot’s knowledge, beliefs, and assumptions. Therefore, the robot should use explicit verbal statements and questions so that the child can build common ground, in addition to implicitly contributing to common ground through gaze behaviours.

For example, the robot should firstly seek explicit alignment on various aspects of the learning assignment: a) alignment on the learning objectives (e.g. “I think the idea of the assignment is to investigate topic A” or “What do you think topic A is about?”); b) the steps to take towards achieving these learning objectives (e.g. “To investigate topic A we need to complete tasks X, Y, and Z” or “What information from task X do we need to use in task Y?”); c) the actions that are required within each step (e.g. “In this step we must first place ‘object 1’ here and then ‘object 2’ there”); and d) the references to physical objects and locations (e.g. “Is this the ‘object’ that we have to use in this task?” or “There is the ‘location’ where we have to put the object”).
Secondly, depending on the available sensors and actuators, the robot may attempt to emulate the gaze behaviours that we observed with children: a) emulate paying attention by reactively gazing towards the child’s actions; b) emulate mutual gaze by reactively gazing towards the child when they gaze towards the robot; and c) emulate shared focus of attention by gazing towards where the child gazes or points.

DG 8: The robot’s actions should be directable by the child and should be communicated as such

According to Klein and Feltovich [107] directability of each partner’s actions plays a key role in successful interdependent activities. Through mutual directability of their actions, collaborators signal to each other that their behaviours affect one another in a meaningful way. In observations of children, we saw instances where children would mutually direct their actions, or where one child would overrule the actions of the other.

We imagine it can be challenging to design robot behaviours in such a way that they are truly directable by the child throughout an entire interaction. To exhibit some level of directability we suggest that designers focus on implementing directable actions that are isolated to specific aspects or moments of an interaction.

For example, asking the child for help (e.g. “Can you tell/show me how to do this?” or “What do I have to do next?”), explicitly asking for directions (e.g. “Which object would you like me to use?” or “Where would you like me to place this object?”), inviting the child to correct the robot’s actions (e.g. “Am I doing this right?” or “Which answer should I pick?”), or synchronising timing of actions (e.g. “Shall I do this now?” or “Let’s count down together!”). In addition to explicit actions the robot could use implicit social cues to signal that its behaviours are directable. For instance by displaying hesitation or waiting for mutual eye contact before executing an action, or by allowing the child to interrupt the robot’s speech and behaviours.

DG 9: The robot should use appropriate signalling behaviours for the coordination of joint activities

Finally, Klein and Feltovich [107] state that to effectively plan and execute joint actions collaborators must clearly signal their own intentions, goals, and actions. In addition they must signal their recognition and interpretation of their partner’s intentions, goals, and actions. Clear and legible signalling is particularly important in situations where common ground is weak (e.g. during first familiarisation), where something in the joint task has gone wrong (e.g. misunderstood task instructions), or when partners’ goals or actions have become misaligned (e.g. working on different sub-tasks).

When using digital learning environments or sensorised physical learning materials, the system is capable of reliably recognising the user’s actions, enabling the robot to detect and respond to deviations from task instructions. However, reliable automatic detection, recognition, and interpretation of an interaction partner's underlying intentions and goals is quite challenging, which makes it all the more important for the robot to use appropriate signalling behaviours towards the child. The robot's signalling behaviour can be composed of explicit verbal statements and implicit non-verbal behaviours.

For example, in some cases, explicit verbal statements can be aimed at communicating the robot’s goals and actions (e.g. “I am working on ‘goal A’ now” or “I think we should first focus on achieving ‘goal A’, then move onto ‘goal B’ afterwards” or “I’m going to do ‘X’ now” or “I’d like to do ‘X’, can you help me by doing ‘Y?’”). In other
cases, the robot may explicitly ask the child to clarify their goals and actions (e.g. “Are you working towards ‘goal B’?” or “Are you going to do ‘Y’ next?”). Explicit verbal statements of intention can be supported through implicit non-verbal behaviours. For example, through attentive gaze, mutual gaze, and shared focus of attention the robot signals to the child that it is “following”, “paying attention to”, and attempting to “recognise” their actions. Furthermore, through proactive behaviours and actions (e.g. by proactively handing over objects to the child, or by reciting information relevant to the task at hand) the robot signals that it is has interpreted the child’s intentions and is anticipating their next action accordingly. Additionally, such proactive behaviours may be signals of the robot’s benevolent intentions.

Recommendation 3 - Designing relational behaviours: The robot should work towards building and maintaining a positive social relationship with the child

Effective collaborative learning requires a positive social relationship between learning partners. Partners need to judge each other’s interests, abilities, skills, and knowledge, and must trust each other to act benevolently. Learning partners can take advantage of their social relationship to offer each other positive emotional support when faced with new experiences or setbacks.

The children who participated in the contextual analysis were from the same class and had already established interpersonal social relationships. Although we did not observe the formation of new social relationships, we did observe instances where children engaged in meaningful relational interactions. For example, children had mutual emotional experiences in response to successes and failures, they displayed prosocial behaviours towards each other, and they engaged in entertaining activities outside of the learning assignment. Designers of learning interventions using robots and other social agents, can take inspiration from such observations. Previous research has shown that children are prone to build relationships with robots (e.g. [13, 94, 96, 202]). Fostering positive child-robot relationships can indeed have beneficial effects for the child (e.g. [10, 98, 203]). The following design guidelines are suggestions on how to implement relational behaviours on an educational social robot.

**DG 10: The robot could display emotional responses to learning events**

Children often display an emotional response when encountering learning events such as successes and failures. By designing the robot to display similar emotional responses it can attempt to emulate a mutual experience with the child.

For example, a basic version of such a robot may use pre-scripted responses to successes and achievements (e.g. “Yeah! We got it right!” or “Aha! I knew it!” or simply cheering and/or smiling). More advanced technological implementations may attempt to analyse the child’s facial expressions, body poses, gestures, and affective speech to offer a tailored response, more accurately emulating a mutual emotional experience. If the intervention and the context of the learning assignment are appropriate, the robot may also emulate mutual experiences during failures (e.g. “Hrmph. Now I have to start all over again!” or “Grrr! I was so sure my answer would be correct!”). However, a negative response may also have a detrimental effect on the child’s learning experience and motivation.
DG 11: The robot could offer emotional support

In addition to mutual experiences, there are several opportunities for the robot to offer more explicit emotional support when children encounter frustrating setbacks that may leave them demotivated.

For example, firstly, while children work on a task, the robot may offer friendly motivational support, emphasising their task-related effort (e.g. “I see you’re really doing your best!” or “You must have thought really hard on that answer!”). Secondly, the robot could focus on offering positive feedback when children experience successes (e.g. “Well done! Onto the next challenge!”) as well as when they encounter failures (e.g. “That was a good effort, just a bit more practice and I’m sure we will make it!” or “I don’t mind making mistakes, as I always learn something new!”). Finally, as children advance through various levels, assignments, or difficulties, the robot could offer praise regarding their progress (e.g. “Great! Thanks to your hard work we got a new highscore!” or “We really nailed that task, well done!” or “I see you learned a lot, let’s try an even more difficult challenge next!”).

DG 12: The robot could show prosocial behaviour

In the contextual analysis we found instances of reciprocal, prosocial, benevolent behaviours between co-learners. In cases where it contributes to the learning experience, a designer may choose to translate specific instances of prosocial behaviours to the robot.

For example, firstly, the robot could exhibit proactive behaviours that contribute to task progress (e.g. the robot offers help at appropriate moments, the robot hands over objects, fills out answers, or reads relevant task instructions). Secondly, the robot may offer the child the opportunity to perform the most amusing, comical, or entertaining actions themselves. What these actions are depends to a large extent on the task context; for our inquiry learning assignment children seemed to mostly enjoy manipulating the physical objects when preparing (e.g. placing pots on the balance) and initiating the actual experiment (e.g. removing wooden blocks from under the balance). Finally, on occasion the robot may directly offer the correct answer when the child is stuck (e.g. “Psst..I think the answer should be ‘A!’”), thus enabling them to continue with the remainder of the assignment. However, it is likely that directly prompting such answers doesn't contribute much to the child's knowledge. Therefore, we advise that the robot only uses this particular relational prosocial behaviour if the answer at hand is not directly related to the main learning goals of the intervention. Otherwise, if the answer is directly related to the learning goals, the robot should offer educational scaffolding instead.

DG 13: The robot could engage with the child in activities beyond learning

We observed many positive relational interactions (e.g. mutual smiling and laughing) when children engaged in non-learning activities (e.g. neatly stacking papers or participating in mischief). More generally, past research has shown that engaging in off-topic activities with a social agent, such as social dialogues and smalltalk, can contribute to a positive relationship [19, 18].

While we do not necessarily recommend that the robot promotes mischief, it still might be beneficial to the social relationship if the child and robot engage in non-
learning activities. In practice, such activities would depend on the child’s personal interests and the robot’s (conversational) capabilities.

Examples of activities that children might like to do with a robot could include drawing and colouring, dancing, storytelling, or playing games.

**Recommendation 4 - The robot should stimulate the cognitive processes of inquiry learning**

In effective inquiry learning, students alternate between processes that are practical in nature (e.g. the process of preparing, experimenting, and observing) and cognitive in nature (e.g. the process of hypothesising and concluding) [104]. Learning interventions should focus on promoting an optimal balance between both types of processes. However, in the observations we found children’s practical actions to be disproportionally more prevalent than behaviours associated with cognitive processes. Designers of learning interventions may address this by offering scaffolding to trigger more cognitive behaviours.

**DG 14: The robot could attempt to elicit explanations from the child**

Cognitive thought processes take place internally in the mind of the learner and can be difficult to observe externally. However, we observed that children would verbally express aspects of their reasoning when explaining their hypotheses and conclusions to each other, offering us a glimpse into their thought processes.

To promote more of such cognitive processes the robot could attempt to elicit explanations from the child. Such an approach could be similar to the interactive explanation paradigm [168] outlined above in DG 3.

For example, the robot could ask the child to further clarify their hypotheses (e.g. “Why do you think that X will happen?”), explain the state of the learning materials (e.g. “Can you explain why the balance tilted this way?”), or formulate underlying physics principles in their own words (e.g. “How do you know which pot is heavier?” or “What happens when pots move closer/further from the pivot?”).

### 4.6 Limitations and future work

In the contextual analysis we observed 22 children learning together in pairs in their usual school environment. Due to the relatively low sample size we may not have gathered an exhaustive collection of all possible educational, collaborative, and relational interactions. However, we did observe sufficiently rich and diverse behaviours for the purpose of specifying the set of relevant design guidelines.

Since the children participating in the contextual analysis all came from a single Dutch school we cannot easily generalise these results to different cultures. Moreover, the participating school followed the Montessori educational system, in which children are quite accustomed to collaborating together on embodied learning tasks. In particular, more knowledgeable older children are regularly paired up with their less knowledgeable younger classmates. As a result, we may have observed different behaviours if the contextual analysis was conducted in schools following different educational systems. However, in later stages of the research, implemented variants

---

1DG 3: A robot operating in the role of differently knowledgeable or less knowledgeable other could leverage the interactive explanation paradigm
of the CAL system were tested with children from a wider selection of schools and daycare centres. We did not observe any behaviours with these children that would point to these design guidelines being biased towards the Montessori system.

The design guidelines presented in this chapter followed from observations of 6-9 year old children. Children at either end of that age group are likely to have differences in their skills and cognitive development. The guidelines do not specifically consider such differences between the younger and older children in an absolute sense. Instead, they offer suggestions from a relative point of view; in terms of a more knowledgeable child with a less knowledgeable partner or vice versa. Furthermore, we cannot generalise the results to children, students, or adult learners that fall outside our age group. In fact, we speculate that more advanced learners working on more complex and challenging assignments would likely benefit from different forms of (social) scaffolding. Therefore, designers of learning interventions may wish to adjust the difficulty of the task or the form of scaffolding to better cater to the needs of children in their target age group.

It is likely that the context and nature of the inquiry learning assignment has inherently shaped the interactions that we observed between children. This type of assignment affords a set of social learning behaviours, that is specific to inquiry learning contexts. We can therefore not generalise the recommendations and design guidelines to other types of learning paradigms.

Future work could focus on collecting and analysing more elaborate and diverse samples. Firstly, the annotation scheme could be further refined through another iteration of inductive coding with a fresh data set. This would allow for more descriptive labels to emerge from salient events that were now labelled as Supplementary interactions, especially on the level of relational interactions. Secondly, including children from different educational systems and cultures could be a first step towards strengthening the empirical basis for the proposed design guidelines and would allow for more generalised applications.

### 4.7 Conclusion

In this chapter we analysed how children worked together on a learning task. We used the ECR model as a lens to look at three types of social learning interactions: 1) Educational interactions, 2) Collaborational interactions, and 3) Relational interactions. We presented a contextual analysis from which we derived recommendations and design guidelines for designing educational social robot interventions. We observed 11 pairs of real learners in a school while they were working on our specific collaborative inquiry learning assignment and learning materials. The video recordings of these sessions were annotated regarding the three ECR types of social interaction. These annotations were analysed to answer the first research question, “Which typical behaviours and social interactions do we identify in learners working with our assignment and how do these behaviours align with theory?” We found that children often behaved as more- or less knowledgeable partners relative to each other while they worked together. In those roles they would scaffold each other’s learning process by thinking aloud, by eliciting or offering explanations, and by agreeing or disagreeing with each other. Furthermore, we found that children coordinated their joint actions by taking turns, discussing goals and plans, and aligning their progress. When doing so, children would often signal their intentions verbally or through gaze and gestures. Misalignment of
actions, goals, or task progress were often explicitly corrected verbally or non-verbally by either child. As they progressed through the assignment, the children showed signs of social bonding by displaying both positive and negative mutual emotions. These observed behaviours were aligned along the three levels of the ECR model.

We also analysed the same data regarding the physical and cognitive behaviours that children expressed in the various phases of the inquiry learning cycle, to answer the second research question, “How do children progress through the various phases of the learning assignment and do the types of learning behaviour they exhibit vary with these phases?” These analyses showed that our learners displayed a rich repertoire of behaviours and social interactions across all three types, but they were predominantly engaged in activities of a practical nature (i.e. preparation, experimentation, and observation), whereas the more cognitive activities (i.e. hypothesis generation and conclusion) were less pronounced.

Then, we looked at the third research question, “What are opportunities for social robot interventions in regard to interaction patterns, specific behaviours and roles, to contribute to the child’s learning in the assignment?”. To answer this, we used the above results to derive four generic high-level recommendations and fourteen concrete design guidelines for when and how a social robot may have a meaningful contribution to the learning process. These design guidelines can be used as a foundation for developing social robots that can navigate the various ECR types of interaction with the purpose of supporting a child’s learning process.

Based on these insights we have a clearer picture of the opportunities for a robot to contribute to children’s learning and we have concrete guidelines regarding the design of social behaviours for such a robot that might work well in our envisioned context. In the next chapter we introduce the research platform for investigating educational child-robot interactions. We highlight the implementation of both the technical setup and the learning materials and discuss the overall methodology for measuring objective and subjective aspects of children’s interactions with the system. As a first step towards showing the potential value of the design guidelines for producing robots capable of supporting children’s inquiry learning, we then discuss four implemented variants of the Robot-Extended Computer Assisted Learning (RECAL) system in which we translated a subset of design guidelines into concrete robot behaviours. Following chapters then describe how we used this platform to investigate specific ways in which a robot may contribute to the learning process.
“OK I think I’ve got it.” Mr Donovan presses some buttons on a tablet and Chucky springs to life.

“Hi kids!” Chucky waves and puts on a big smile. “I can’t wait to get started! But first I need your help.” He looks towards Mees and Sophie, “I saw you had a task with floating and sinking that looked like a lot of fun, but I’m not a big fan of all that water...” “Do we have any other cool tasks here in class?” he asks, looking towards the large cupboard full of labelled boxes.

Kids rush to their feet and fetch their favourite boxes. “Hmm, ‘Balance’, sounds interesting, let’s start with that!” Chucky says while looking through the box. “I don’t know this task yet, why don’t you all get started setting up while I rework my programming and run some quick updates.”

“I’ll be back!” he says, while powering down his circuits.

Meanwhile, the teacher puts on an energetic tune and cranks up the volume as the children get to work assembling their learning tasks.

5.1 Introduction

In the previous chapter we composed a set of recommendations and concrete design guidelines for educational robots. In the remainder of this thesis we investigate how such a robot may contribute to the learning process of children. To do so, we conceptualised and constructed a research platform that allowed us to run a variety of educational child-robot interaction studies. The platform presented in this chapter incorporates several components: 1) pedagogically appropriate learning tasks; 2) a modular architecture for orchestrating the educational interaction between child and

The platform presented in this chapter materialised over several iterations, parts of which were previously presented in the following publications: Davison et al. [49], Davison et al. [48], Davison et al. [50], Charisi, Davison, et al. [35], Wijnen, Davison, et al. [220], and Reidsma, Charisi, Davison, et al. [171].

Cue the A-Team theme song: https://www.youtube.com/watch?v=wyz_2DEah4o
robot; and 3) measurement instruments for gaining insights into how a robot may affect the child’s experiences and learning. The platform presented here has materialised over several iterations; in the sections below we present the most complete rendition. Furthermore, we discuss how we geared the platform towards interactive learning experiences that are robust during longitudinal in-the-wild deployments. Finally, we briefly discuss four studies using concrete intermediate variants of the platform that were implemented according to the design guidelines.

5.2 Learning tasks

The learning tasks were developed according to a constructivist learning approach [66] to support children’s natural development [167]. We adopted the process of inquiry learning to structure the way in which children explored the learning materials. The following steps of the inquiry cycle were adapted from Klahr and Dunbar [105] and Klahr [104]: prepare, predict, experiment, observe, and conclude. In the prepare phase children would prepare the learning materials according to the assignment instructions. In the predict phase children would be asked to give their hypothesis about what would happen with the learning materials. In the experiment phase children would conduct the actual experiment using the physical learning materials. In the observe phase children would indicate what had happened during the experiment. In the conclude phase, they were asked to reflect whether their initial hypothesis was correct or incorrect.

The amount of information and guidance offered to the learner influences the complexity of the task. Bell et al. [9] and Banchi et al. [2] describe a continuum of several levels of inquiry. Learners start with highly-structured “confirmation inquiry tasks” in which questions, hypotheses, methods, and results are explicitly provided. They then gradually move to more “open inquiry tasks” in which they formulate their own questions, hypotheses, and methods, after which they construct arguments to explain experiment results [2]. By offering appropriate scaffolds (e.g. predefined task structures or explicit instructions) learners are able to progress from closed to more open forms of inquiry.

Two learning tasks were implemented as part of the platform. Each learning task was centred around a phenomena from the physics domain. Tasks consisted of several difficulty levels which the children could complete at their own pace.

5.2.1 Balance scale task

In the balance scale task children explored the concept of moment of force. This task was a continuation of the one used during the contextual analysis. The materials consisted of a wooden balance scale with three pins on each side, a total of six pots of various weights, and two wooden blocks, as shown in Figure 5.1a. The colour of the pots represented their weight: grey pots were the heaviest, red pots were lighter than the grey pots, and yellow pots were the lightest. Pots could be placed over the pins on the balance at increasing distances from the central pivot. The two wooden blocks could be placed under either side of the balance to prevent it from tipping over when preparing an assignment (i.e. placing or removing the pots and making a prediction). After children had placed the pots and had made their prediction, they removed the blocks to observe what would happen with the balance (either tilt left, right or remain in equilibrium) and draw conclusions regarding their prediction. By
The balance scale task where children place weighted pots on either side of a central pivot.

The ramp task where children race various balls from two angled slopes.

Figure 5.1: Examples of the implemented learning tasks.

The assignments with the balance scale had difficulty levels ranging from 1 to 8. Levels 1 to 5 followed a structured inquiry approach as described by Bell et al. [9]: the assignment stated the selection of pots that were to be used and the location in which they should be placed. The exact exercise at each level was generated at random according to the following rules. In level 1, two pots of the same colour were used. These were placed on the same distance from the pivot on both arms of the balance. The result was that the balance always remained horizontal. In level 2, again two pots of the same colour were used. The pots were now placed on different distances from the pivot. The result was that the balance either tilted to the left or right. In level 3, two pots of different colours were used. The pots were placed at the same distance from the pivot. Due to the difference in weight, the balance either tilted to the left or right. In level 4, both variables were manipulated at the same time: two pots of different colours were used and the pots were placed on different distances from the pivot. As a result the balance either tilted to the left, to the right, or remained horizontal. In level 5, three pots were used; two of the same colour and one other colour. The pots of the same colour were placed together on the left side of the balance. The other colour was placed on the right side. The result was that the balance either tilted to the left, right or remained horizontal.

In levels 6 to 8 the assignments progressed towards a more guided inquiry approach: the assignment stated which pots to use, but the children chose where to place them. In level 6, children used two different coloured pots. They were asked to place one pot on a given pin and to decide for themselves where to place the other pot to keep the balance horizontal. In level 7, children used two pots of different colours. The children were asked to place both pots in such a way that the balance remained horizontal. In level 8, children used three different coloured pots. They were asked to place the pots on the balance in such a way that the balance remained horizontal.

5.2.2 Ramp task

The ramp task involved the exploration of potential energy and rolling resistance. Children used a ramp with two angled slopes from which they could race balls of different materials, weights and sizes, as shown in Figure 5.1b. They explored how the angle of the slope and the properties of the balls affected the time it took each ball
to reach the end of the ramp. The following balls were used: two small marbles, two
table tennis balls, two rubber beach tennis balls, two large marbles, two wooden balls,
and two styrofoam balls. The slopes could be individually adjusted in a low angle
(approx. 10 degrees) or a high angle (approx. 20 degrees). The balls would be placed
on a fixed point at the top of the slope. The children could press a button to initiate a
short countdown, after which both balls would simultaneously be released and start to
roll towards the finish.

The ramp assignments had difficulty levels ranging from 1 to 9. Similar to the
balance scale task, level 1 to 6 followed a structured inquiry approach: the assignment
stated the angle at which to set the slopes and which balls to use. The exact exercise
at each level was generated at random according to the following rules. In level 1 the
two slopes were positioned at an equal angle. Two of the same types of balls (e.g. two
small marbles) were placed and rolled off the ramp simultaneously. In level 2 one
of the slopes was angled higher than the other. Two of the same balls were placed
and rolled off the ramp simultaneously. In level 3 both slopes were again angled
equally. Now, two different balls (e.g. a table tennis ball and a styrofoam ball) were
placed and rolled off the ramp simultaneously. In level 4 one of the slopes was angled
higher than the other. Again, two different balls were placed and rolled off the ramp
simultaneously. In level 5 both slopes were angled equally high. Now, three different
balls were placed, where one was placed behind the other ball (e.g. a wooden ball is
placed behind a styrofoam ball and on the other ramp a small marble is placed). The
balls rolled off the ramp simultaneously. In level 6 one of the ramps was angled higher
than the other. Again, three balls were placed where one was placed behind the other
ball. The balls rolled off the ramp simultaneously.

In levels 7 to 9 the assignments progressed towards a more guided inquiry approach,
similar to the balance task: the assignment stated an objective, and the children could
use their own approach at finding an answer. In level 7, the objective was to find out
which balls were faster or slower than a certain ball. For example, “find all the balls
that are faster than the table tennis ball”. Children could repeat as many experiments
as they needed. In level 8 children were instructed to use the stopwatch on the tablet,
and could only use one slope at a time to find the fastest or slowest of two balls. They
had to repeat the experiment several times, and remember/write down the times
between experiments. This was a preparation exercise for the final difficulty level. In
level 9 children had to make a ranking of all the balls ranging from fast to slow, or
slow to fast. Again, they could repeat as many experiments as they wished.

5.3 A modular architecture for Robot-Extended Computer Assisted Learning (RECAL) systems

The modular architecture originally developed as part of the EASEL project is used as
a foundation for the RECAL system [171, 211]. The architecture is ideally suited for
supporting fluent multimodal interactions [49] and longitudinal interactions across
multiple sessions [48]. Conceptually, the architecture is organised according to three
layers: perception, interaction management, and behaviour realisation. Several system
components and software modules work together across the layers to orchestrate the
educational interaction, shown in Figure 5.3. Components communicate and exchange
data through a common middleware interface, allowing a flexible and extensible setup.
The platform supports TCP/IP, UDP, ActiveMQ, YARP, ROS, and Apollo. To integrate
with external libraries, individual modules can be configured to use any combination of these middlewares. Figure 5.2 depicts a typical setup of the RECAL system.

5.3.1 Perception

The perception layer is responsible for informing the RECAL system about the state of the world, relevant to the interaction. Individual modules sense and interpret aspects about the user and the user’s actions:

**User detection and tracking** A Microsoft Kinect sensor running the SceneAnalyzer software [232] detects a child’s presence and their location relative to the robot.

**User identification** Individual children are identified by scanning their personal RFID card at the start of an interaction session.

**Action recognition** Embedded sensors and microcontrollers in the learning materials recognise in-task actions. The balance scale task contains the following sensors:

- Resistors are embedded in the base of each pot. By measuring the resistance on each pin of the balance, the system detects which pots are placed on which position
- Magnets are embedded in the top and bottom of the wooden blocks. Reed switches are used to detect whether the blocks are placed under the balance
- A potentiometer in the pivot point measures the tilt of the balance

The ramp task contains the following sensors:

- An optic sensor at the end of the ramp measures the time between the ball release and their arrival at the end.
- A potentiometer in each slope measures the tilt angle
User input  The primary means of capturing user input is through multiple-choice buttons on a tablet interface and through voice activity detection. Interpreted data from these perception components is forwarded to the interaction management modules.

5.3.2 Interaction management

The interaction management layer orchestrates the flow of the interaction between child and robot, and their progression through the learning task. Perception data is fed into the dialogue engine Flipper [139, 215]. Flipper is an Information-State Update engine that consists of collections of templates that work together to generate appropriate responses to external events. A template consists of preconditions, effects, and optional behaviours. Preconditions are rules that are checked against the information state. The effects are executed if all preconditions of a template are true. Effects update the information state, possibly triggering a cascade of other templates. For the interaction manager we defined three models to manage various aspects of the educational interaction:

**Memory** Persistently stores and retrieves information about individual children across multiple interactions based on user identification data.

**Inquiry phases** Models the abstract phases of the inquiry learning cycle as a finite state machine. Transitions between states are based on action recognition and user input data.
Instruction and feedback Models a concrete instance of a task. Assignments are generated based on the child’s interaction history loaded from memory. The model generates step-by-step instructions for the assignment based on the current phase of the inquiry cycle. Feedback, scaffolding, and personalised praise is generated and triggered based on action recognition and user input data.

Whenever a dialogue engine template in one of the models is executed, any associated behaviours are forwarded to the behaviour realisation modules.

5.3.3 Behaviour realisation

The behaviour realisation layer takes care of scheduling, planning, and executing the behaviour of the RECAL system’s embodied components: the robot and the tablet. Behaviours originating from the interaction manager are specified in generic Behaviour Markup Language\(^1\) (BML) \([110]\) which is translated by ASAPRealizer \([172, 218]\) into commands that are specific to a certain embodiment. By default, ASAPRealizer supports gestures, animations, and speech behaviours (prerecorded and text-to-speech). Exact timings can be synchronised to ensure coherent and logical behaviour sequences. ASAPRealizer was further extended to support the following embodiments:

**Robot** RoboKind’s R25 Zeno\(^2\) robot, as shown in Figure 5.4, supports facial expressions, gaze shifts, and upper-body gestures. Gaze shifts can be predefined static locations (e.g. the learning task) or dynamic locations (e.g. the user’s face, derived from user detection and tracking data). ASAPRealizer generates lip-synchronisation based on the system’s speech.

**Tablet** The system’s Android tablet app supports multimodal content presentation, as shown in Figure 5.5. It displays task instructions using images and text and captures user input through multiple-choice buttons.

Together, the three-layered modular approach of the RECAL system enabled us to construct a variety of learning setups. Depending on the study, we could choose

---

\(^1\)Behaviour Markup Language: [http://www.mindmakers.org/projects/bml-1-0/wiki](http://www.mindmakers.org/projects/bml-1-0/wiki)

\(^2\)Robokind: [https://www.robokind.com/](https://www.robokind.com/)
to implement and combine individual components, each with varying degrees of complexity and autonomy. Four concrete studies are discussed in Section 5.6 of this chapter.

5.4 Methods and measures

Conducting studies with young children necessitates measures that are appropriate for the target user group. In a recent systematic review of evaluation methods in child-robot interactions, Charisi et al. [33] underlined the need for new methodological approaches to measure how children work with and respond to educational robots. To investigate how the robot affected children’s learning process we relied on a collection of quantitative and qualitative measures. Using an iterative approach we constructed a collection of measurement instruments consisting of subjective measures, such as observations, (semi-structured) interviews, and self-reports; and objective measures, such as task usage patterns. Together they offered us insights in how children interacted with the robot and task, what their impressions were of the robot and task, their attitude towards learning, how they appraised and reflected on their own learning process, and how they self-regulated their learning process. Here, we

---

1Variants of this architecture have since been used in other projects: in De-Enigma [38] it is used to control a Zeno robot that practices facial features and expressions with autistic children; in COUCH [196] the architecture is used to manage a council of virtual coaches that help older adults achieve healthy living goals; in R3DS [138] the architecture controls the behaviour of a robot & virtual agent receptionist duo; and in Snoozle [212] it animates a non-anthropomorphic robotic pillow that entices users to maintain healthy sleeping habits.
briefly present each instrument and how it materialised; more specific details regarding their implementation and usage are given in later chapters.

**Observational measures of verbal interaction** Similar to the approach taken during the contextual analysis we used observational measures to analyse verbal interactions between children and the robot. Based on a grounded theory approach, we used axial coding to derive a coding scheme for verbal explanations that children gave to the system. The coding scheme covered the breadth (the number of mentioned topics) and depth (the number of associations drawn between topics) of their explanations. Further details are discussed in Chapter 6, where we investigate how children explain their reasoning process to a robot.

**Semi-structured interviews to measure perceived sociality** Over the course of several iterations we constructed two instruments that used (semi-structured) interviews. The first measured aspects regarding children's perceptions of the system's social competence. It was developed in two iterations and validated through divergent and convergent validity [35, 50]. Children were asked open questions regarding their thoughts on the social aspects of the system. During interviews we loosely followed a script, leaving ample room to further expand on specific lines of thought. Additionally, we used pictorial exercises as a way to focus the conversation on certain social topics familiar to the children. To evoke meaningful and more elaborate responses we often asked children to further explain why they gave a certain answer. A coding scheme was derived through axial coding. The coding categories covered task-related elements, aspects of the system's design or technology, the role of the system, interactions with the system, theory of mind ascribed to the system, and the system as a machine / social artefact / social agent. In Chapter 6 we further discuss how the interviews were used to investigate children's views of the robot's social competence.

**Semi-structured interviews to measure mindset in learning** The second interview instrument measured children's mindset towards their own intelligence and learning process and was originally developed by colleagues from educational psychology as a questionnaire for older children. Questionnaire items were adapted for younger children and were administered through an interview. The interview method was iteratively designed and tested during several pilot studies. Children indicated their level of (dis)agreement with statements that were read out loud by the researcher. To help children nuance their answer to each statement, children were first asked whether they agreed or disagreed and were subsequently asked whether they (dis)agreed fully or partially. These interviews are used in Chapter 7 to measure changes in children's mindset while they interacted with the system for an extended period of time.

**Self-reports embedded in the task to measure experiences** We developed unobtrusive in-task self-report measures to gain insights into how children's experiences and attitudes developed over multiple interactions with the system. Multiple-choice questions were embedded as part of each assignment. Since these used the same system speech and tablet interface they did not interfere with children's regular routine. The questions covered topics of task performance and appraisal thereof, experienced task difficulty, and exerted effort. These measures were used in Chapter 8 to follow the development of these aspects over time.
**Usage patterns** To follow how children’s usage patterns developed over time we automatically collected logbooks of their interactions with the system. The logbooks contained information regarding interaction frequency and duration, and their progression through levels of the task. Chapter 8 explains these measures in more detail.

5.5 Gearing the platform towards sustaining long-term interactions

Besides single-session studies, we also carried out a longitudinal study to see how children’s interactions with an educational robot developed over multiple interactions. Investigating this required that we design tasks and interactive system in such a way that they could be deployed longitudinally in ecologically valid contexts familiar to the children. For this purpose, a system must function reliably and autonomously without supervision from a researcher [6].

Longitudinal studies, especially those with repeated interactions, require sufficiently rich and varied content [43]. In general, it is important that there is an appropriate amount of content to keep the target group engaged over extended periods of time. Educational content in particular should cover topics that align with the target user’s interests and should be geared towards their abilities. Since their abilities and interests might change during the course of longitudinal interaction, it is important that the content and the nature of the activities can accommodate such changes. For instance, educational activities for children should be designed to include adequate difficulty levels to cater to a wide variety of children at various stages of development. Children should remain engaged and challenged as the study progresses and they acquire more skills and knowledge in the subject domain. To avoid overly repetitive interactions designers may consider adding a level of variation to the system’s content. In educational activities, certain aspects and parameters of the learning task can often be varied without influencing the task difficulty. Moreover, at various points in an interaction different variances of speech phrases could be used to deliver the same message to the user. Both of the learning tasks were designed to address these points.

Socially intelligent interaction systems operating in the wild must deal with very diverse users in a wide range of contexts. Designers of these systems often have to consider certain trade-offs to offer the users an intuitive, compelling experience while maintaining the robustness of the interaction [177]. The core interactions with the system always revolved around the learning materials (i.e. placing pots and rolling balls), which were progressively enhanced with interactive instructions and feedback. To ensure that the core interaction could always take place we designed interaction patterns that were robust against various technical and interactional breakdowns. Where possible, we followed advice from Leite et al. [128] in the design of the interactive system. In their survey and underlying discussion of the included papers they propose several generic guidelines for long-term human-robot interactions. Here, we discuss the rationale behind the design of the system, aimed at supporting robust interactions in the context of our learning tasks.

5.5.1 Incorporating robust interaction patterns

Designing a compelling interactive system can be a delicate balancing act. On the one hand, interaction designers want to ensure that users can interact with their system as
naturally, intuitively, and fluently as possible. This can be achieved through state-of-the-art perception and reasoning modules, specialised in detecting, recognising, and interpreting human (social) behaviour (e.g. the user’s speech, their facial expressions, gaze, body posture, gestures, and so on). Many of these perception and reasoning modules perform well in controlled setups. However, environmental factors (e.g. noise or lighting conditions) and human factors (e.g. age, gender, appearance, or cultural differences) may influence their accuracy and reliability when deployed in less controlled real world settings. Such detection and reasoning errors may then lead to miscommunications, unpredictable or illegible responses from the system, breakdowns in the interaction, and frustration from the user.

On the other hand, interactive systems should ideally be designed to offer the user a robust and reliable interaction to prevent such errors from occurring in the first place. Designers could take this into account by incorporating interaction input interfaces that are less error-prone to outside factors. For example, many users will be familiar and proficient with more traditional interfaces, such as mouse, keyboard, and (touch)screens. Incorporating such interfaces may increase the robustness of the interaction in situations where automatic perception and reasoning systems might struggle, and may offer the user a more reliable and direct interaction interface.

To ensure a robust educational experience we excluded as many potentially unreliable interaction methods as possible. Therefore, we designed interactions with the system to primarily incorporate explicit user input through a tablet touchscreen and implicit input through manipulations of the learning materials.

In addition to reliably capturing the user’s actions, the system’s responses might not always be noticed, heard, or understood by the users. This might be the case when users are distracted, when there is too much environmental noise, or simply because the interactions are unfamiliar to them. To avoid and recover from these situations, designers could use implicit and explicit dialogue design patterns. Firstly, for most dialogue moves designers could consider adding one or several repetitions, that trigger automatically when there is no response from the user after a certain timeout. Secondly, in combination with repetition designers may implement multiple similar variations of each dialogue move where appropriate, such that the user stands the best chance of correctly interpreting and understanding the system’s intents. In our system such variations and repetitions were implemented for each phase of the learning task. Thirdly, designers could incorporate alternate dialogue paths that may either accomplish similar interaction goals in different ways, or may enable the system to respond to more varied dialogue moves or actions from the user. This was implemented in cases where children skipped a step of the inquiry cycle. Children could continue with the remainder of the task, but the system would then remind them of what they skipped. Finally, to cater to a wider audience of users who might have specific preferences or impairments, or are on different levels of development, the system may incorporate fallback options in its responses. Our system offered multimodal responses and task instructions (i.e. through text, images, movement, and sound). Additionally, children could request elaborated explanations at various phases of the task.

Finally, if we wish to deploy socially interactive systems in unsupervised settings (i.e. without a researcher present) they must be able to autonomously interact with users and transition between sessions. Although we cannot account for all unforeseen events, the dialogue should at least be designed to support various contingencies
where users initiate and discontinue interactions. For example, consider the simple case where a user starts the interaction and then walks away after a few moments. In this case, the system may recognize this and halt the current interaction after a timeout duration, then finally it could reset itself to be ready for a next user. Implementation of such autonomous behaviors requires some form of detection of user presence and user identification. In our case we implemented user identification through unique RFID tags. User presence was determined based on a visual system and on children’s interactions with the tablet interface and the physical learning materials. After several minutes of inactivity the system would assume the child had left and the system would automatically reset to an initial state.

5.5.2 Incorporating robust technical systems

Interactive social systems often consist of many different technical components and modules that work in sync to support the ongoing interaction (e.g. see [28, 29, 62, 208, 211]). Typically, there will be modules to detect and interpret the user’s actions, reason about and determine the appropriate system response, and finally plan, schedule, and execute the exact actions associated with that response. Unfortunately these systems can become quite technically complex, with many different types of software running on different platforms and physical machines. Some modules can be considered stable production-ready software, while others may be alpha versions of experimental research software. In our experience, when using complex systems for long periods of time, it is not uncommon for individual components to crash, run into memory problems, report incorrect results, or for communication between components to malfunction. In unsupervised deployments in the wild there might not be a researcher or technician present on-site to notice, troubleshoot, and resolve such errors as they occur.

To ensure that the system was capable of offering a robust interaction, even when some individual components may have failed, we adopted a graceful degradation and progressive enhancement approach. The terms graceful degradation and progressive enhancement have long been used in different fields, like web development¹. There, the terms are used to describe the process of respectively removing or adding (non-essential) functionality and features to webpages in such a way that the core interaction is maintained even when a user’s browser does not support certain technologies. Applying these design approaches to complex interactive social systems such as robots requires some re-interpretation:

1. Using graceful degradation a fully-functional interactive system gradually reduces its functionality when certain components fail or become unavailable. The system architect and dialogue designer should consider how the interaction with the user changes when an individual non-essential component can no longer be used. When that happens the system may degrade it’s functionality by adjusting the dialogues and interactions to no further rely on that component. In these cases the system could inform the user about the failures, which might help them to adjust to the new situation. In practice this may mean falling back on alternatives (e.g. in case a speech recognition module should fail: “I think my hearing is broken, you may use the buttons on the touchscreen instead”) or

¹Example of graceful degradation and progressive enhancement in web development: https://www.w3.org/wiki/Graceful_degradation_versus_progressive_enhancement
default values (e.g. when an embodied agent’s gaze tracking module fails, the agent could instead use predefined static gaze targets), depending on the actual system at hand.

2. Using progressive enhancement a basic stable “core” set of key interactions is gradually extended with more advanced functionality when certain components are available. At its core the system should offer only the essential interactions that are absolutely required within the given interaction context. Ideally, these core interactions rely on as few auxiliary modules as possible to keep the technical complexity low. These core interactions can then progressively be enhanced with more advanced technological modules, which might enable more dynamic and intricate interactions with the system. For example, a social conversational system at its core could be built around a chatbot-style textual dialogue. All essential core interactions could take place through a keyboard and regular computer screen. Interactions with this system could then progressively be enhanced in several ways, such as: adding an embodied virtual or physical agent, adding speech recognition, adding speech synthesis, adding user facial detection/identification, adding valence and arousal detection, and so on. However, even when any one of these components might stop working the system is always capable of maintaining the core interaction.

In the RECAL system we used a combination of these two approaches to create a robust interaction platform. The educational core of the system is the physical learning task. Even when disconnected from all other systems, children can still work with it as an analogue learning tool. The core of the interaction takes place through the tablet interface. It displays multimodal instructions and feedback and offers a robust way for children to give input to the system. Additional modules were designed to progressively enhance the interactive experience. For instance, speech was played through headphones as an additional modality and the system could display reactive and deictic gaze behaviours through the robot. However, the robot’s gaze generation relied on a module that detected the location of the child’s face. The robot used this location data to dynamically gaze at the user at appropriate moments in the interaction. If the facial data was unavailable during the interaction (due to a crashed module, or otherwise), the robot would gracefully degrade its gaze behaviour to use predefined static gaze targets instead.

Components may fail at any time during an interaction. The system should attempt to recover or restart a component when it has detected that it has crashed, or is otherwise showing problems like memory leaks or reporting invalid sensor data. However, depending on numerous factors (e.g. the system architecture and the nature of the failure) such a recovery attempt might have an impact on the ongoing interaction with the user. For example, restarting a speech synthesis component might influence the system’s output modality (e.g. from text-based to auditory), whereas restarting an agent’s or robot’s control libraries might influence their animations and dynamic movements (e.g. halting gesturing or gaze tracking). Therefore, to prevent distractions or disturbances the system should ensure that recovery attempts occur at appropriate moments in the interaction, a concept which we refer to as “interaction-driven error recovery”.

The exact timings of recovery events will depend primarily on the context of the interaction. In general, a safe moment to perform recovery would be after one user
session ends and before the next session begins. However, in some interactions it might be possible to attempt recovery during an interaction session. For example, when the user is temporarily engaged with an auxiliary task and there is no interaction required with the system (e.g. user is reading a long piece of text), or when there is a transition in the context of the interaction (e.g. user has completed activity A and now continues to activity B). In the RECAL system we used the following approach to restart modules, if necessary: 1) the facial tracking module was restarted between robot behaviours, 2) the connection to the robot control software was restarted between activities, and 3) modules prone to memory leaks were restarted between each session.

5.6 Implemented variants

In this section we briefly discuss four variants of the interactive RECAL system. These variants were implemented and used in various studies, some of which are also in this thesis [50, 52, 51, 163, 184, 220]. We highlight how the recommendations and design guidelines from Chapter 4 manifested in each variant. Furthermore, to gain a first impression of how well we succeeded in translating the design guidelines to actual robot behaviours we briefly reflect on how children interacted (socially) with these variants during empirical studies. This allows us to take a first step towards showing the potential value of the design guidelines for producing robots that are capable of supporting children's inquiry learning. Table 5.1 shows the specific design guidelines implemented in each variant of the system.

Additionally, we discuss which tasks, technical components, and measures were used in each variant. For each technical component we indicate in Table 5.2 whether it was fully autonomous or relied on Wizard of Oz (WoZ) input and whether it used preconfigured static content or dynamically generated content.

5.6.1 Variant I: A robot that elicits and stimulates explanatory behaviours

This variant was published as Wijnen et al. [220] and is discussed in more detail in Chapter 6 of this thesis.

During the contextual analysis we observed how pairs of children progressed through the inquiry learning assignment. We found that children more frequently displayed practical actions (e.g. placing and moving pots) than behaviours associated with cognitive learning processes (e.g. discussing observations and explaining conclusions to each other). Therefore, we set out to implement a RECAL system to stimulate the cognitive inquiry learning processes of hypothesising and concluding. The main rationale was to design a social robot that offered social facilitation to leverage the interactive explanation paradigm [168].

The following primary design choices were implemented to support this paradigm. We explicitly presented the robot to take on a certain role in relation to the child. On the one hand, the robot acted as a less knowledgeable other regarding the principles of balance: 1) the robot was given a background story and was introduced as being curious to investigate the concept of balance together with the child (i.e. the robot offered no explicit content scaffolding); 2) it asked open questions at appropriate

1Recommendation 4 - The robot should stimulate the cognitive processes of inquiry learning
2DG 2: A robot should explicitly take on a certain role
3DG 3: A robot operating in the role of differently knowledgeable or less knowledgeable other could leverage the interactive explanation paradigm
Table 5.1: An overview of the specific design guidelines (DGs) that have been implemented in each of the four RECAL system variants. For each variant we list the primary design guidelines (denoted with a black “✓”) and supplementary design guidelines (denoted with a grey “✓”) that were followed during the design and implementation of robot’s behaviour and the learning assignment. Some design guidelines currently remain without implementation.

<table>
<thead>
<tr>
<th>System variants</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG 1: The robot could use the physical learning materials to supplement its scaffolding</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DG 2: A robot should explicitly take on a certain role</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DG 3: A robot operating in the role of differently knowledgeable or less knowledgeable other could leverage the interactive explanation paradigm</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 4: A robot operating in the role of a more knowledgeable other could offer tutoring to the learner</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DG 5: To express an intention for collaboration, the robot should disclose its own goals and should discuss task objectives, strategies, and joint actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 6: To support a level of interdependence, the robot and child should facilitate each other’s actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 7: The robot should build and maintain common ground with the child in the context of the learning assignment</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DG 8: The robot’s actions should be directable by the child and should be communicated as such</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 9: The robot should use appropriate signalling behaviours for the coordination of joint activities</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>DG 10: The robot could display emotional responses to learning events</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DG 11: The robot could offer emotional support</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DG 12: The robot could show prosocial behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 13: The robot could engage with the child in activities beyond learning</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG 14: The robot could attempt to elicit explanations from the child</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sometimes during the inquiry phases in which the child formulates a hypothesis or draws a conclusion, inviting the child to give an explanation; and 3) it gazed towards the child when asking questions to strengthen the social facilitation (audience effect), inviting the child to give more elaborate and deeper explanations. On the other hand, the robot acted as a more knowledgeable other regarding the inquiry process: 1) as part of the robot’s background story it was introduced as having experience in researching physics phenomena; and 2) it could offer more elaborate explanations regarding the steps of the inquiry cycle (i.e. the robot offered explicit process scaffolding).

In addition, the following supplementary design choices were implemented through-

---

1DG 14: The robot could attempt to elicit explanations from the child
2DG 4: A robot operating in the role of a more knowledgeable other could offer tutoring to the learner
Table 5.2: An overview of the various technical components that have been implemented in each of the four RECAL system variants. For each variant we list whether its modules were operated autonomously (●) or through a WoZ (▲), and whether they used predefined static (■) content or dynamically generated (▲) content.

<table>
<thead>
<tr>
<th>System variants</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User detection and tracking</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User identification</td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action recognition</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User input</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interaction management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inquiry phases</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction and feedback</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behaviour realisation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robot</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tablet</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Out the interaction. Firstly, the robot emulated attentive gaze by following the child’s interactions with the tablet and learning materials\(^1\). Secondly, it emulated mutual gaze and displayed emotions at particular moments in the interaction\(^2\). For example, just when the balance tipped to either side during the experiment phase, the robot would gaze towards the child and smile. Thirdly, the robot used didactic gaze to emphasise aspects of the learning materials\(^3\). For example, it would gaze towards the tablet as new instructions appeared and would briefly gaze towards the balance when explaining or reading task instructions aloud. Finally, the robot offered supportive encouragement and feedback as children progressed through the assignment\(^4\). For example, it praised the child’s progress (e.g. “Well done!”) and it always offered positive feedback without judging the child’s performance (e.g. “That’s interesting!”) when hearing their explanations, regardless of whether they were correct.

Table 5.2 outlines which components of the technical architecture have been included in each variant of the RECAL system. Since the single-session interactions of Variant I were always supervised and initiated by a researcher, no user detection, user identification, and memory modules were included. Action recognition regarding elements of the learning materials was carried out by a WoZ by emulating sensor values (i.e. whether pots were placed in correct positions and whether blocks were placed underneath). Additionally, the WoZ emulated voice activity detection by indicating

\(^1\)DG 7: The robot should build and maintain common ground with the child in the context of the learning assignment
\(^2\)DG 10: The robot could display emotional responses to learning events
\(^3\)DG 1: The robot could use the physical learning materials to supplement its scaffolding
\(^4\)DG 11: The robot could offer emotional support
Research platform

User input from the child's tablet interface was processed autonomously by the system. The system followed the child's progression through the inquiry phases autonomously and the system's static instructions and feedback were triggered autonomously, based on the perception data. However, in rare cases the WoZ triggered a repetition of instructions when the child was stuck (i.e. after a certain duration of inactivity) or when the child asked for help verbally (e.g. “I don’t know what to do”). The RECAL system's assignments, behaviours, and speech were based on static predefined and prerecorded action sequences.

The main study was conducted with children of 5-10 years old and used a combination of observational measures and semi-structured interviews. The study methodology and results are reported in detail in Chapter 6. Here we reflect on the types of behaviours that we anecdotally observed, and discuss them in light of the ECR model.

On the educational level, we observed children giving rich explanations to the robot. This is interesting as it lends potential credence to a less knowledgeable robot as social facilitator to support the interactive explanation paradigm. Additionally, we found indications that children might have appreciated the process support of the robot; they would often request additional help and explanations from the robot when they didn’t understand particular steps of the task, and younger children (who might not yet be proficient readers) often asked the robot to read tasks out loud. These observations were further supported by the exit interviews, during which some children described how the robot helped them during the tasks while others mentioned that they explained the tasks to the robot.

On the collaborational level, we anecdotally observed instances where children asked the robot to give its own input, thoughts, and interpretations to the task (e.g. “What is your opinion about this, robot?”). Additionally, we found that some children seemed to dislike the robot’s repetitive instructions. Due to its pre-scripted dialogue models, the robot would repeat the same instructions across different tasks when scaffolding the inquiry learning process. However, some children may have been annoyed by this repetition, as they stated that they already knew the order of the steps after the first task. Others seemingly took offence that the robot would “warn them” about following task instructions even though they had never once deviated from the task and always followed the instructions meticulously. More adaptive dialogue models could take the child’s past actions into account when offering process support. For example, by only issuing elaborate instructions and reminders when first encountering new steps, or when children repeatedly skip certain steps.

5.6.2 Variant II: A robot that engages in extracurricular activities

This variant was the result of a Bachelor thesis project [184] and was published as Davison et al. [50].

In this variant we were interested in supporting interactions on the relational level\(^1\). From the observations of child-child interactions we found that children displayed mutually positive social emotions when they engaged in extracurricular activities. To stimulate such behaviours when interacting with the robot we implemented an additional activity besides the learning assignment.

\(^1\)Recommendation 3 - Designing relational behaviours: The robot should work towards building and maintaining a positive social relationship with the child
This variant built upon the same robot, interaction, and learning assignment as in Variant I; as a result, it addressed the same design guidelines and used the same technical architecture. The primary design effort in this variant, however, was focused on developing an additional intermediate extracurricular activity\(^1\). During this (hopefully) fun and engaging activity, child and robot attempted to make the same dance-like movements with their arms and legs. This shared experience was expected to influence how children perceive the robot on a social level. The robot would encourage the child to do their best, but would also mention that it is unable to perform all the movements itself.

This exploratory study was conducted with children of 5-11 years old and used a combination of observational methods and exploratory semi-structured interviews with pictorial tasks. The main results are reported elsewhere in Davison et al. [50] and Schindler [184]. Here, we briefly reflect on anecdotal observations of behaviours on the relational level. Firstly, we found indications that children seemed to enjoy the extracurricular activity with the robot, as we annotated more smiling during and after the activity. Secondly, through exploratory semi-structured interviews and picture association tasks we attempted to gather insights regarding how the children perceived the robot on a social level. These highly speculative results seem to suggest that children considered the robot to be a social agent of sorts, of a similar age to themselves. In their descriptions of the robot most children mentioned that they thought it could be their friend, and they would describe hobbies, activities, and games they would like to do with the robot. Most children indicated they would like to sit next to the robot in class and that they would feel comfortable to entrust the robot with a secret. Finally, some children indicated that they were amused that they were able to do things that the robot could not. Others mentioned they were impressed by the robot’s overall skill, in both the learning assignment and the dancing activity.

5.6.3 Variant III: A robot that automatically offers appropriate scaffolding

This variant was the result of a Bachelor thesis project [163], parts of which were published in Davison et al. [48].

Variant I and Variant II used pre-scripted fixed-pattern scaffolds that were partially triggered based on WoZ input. However, to support a fully autonomous interaction, the RECAL system must be able to recognise at which point in time it should offer which type of support to the child. Therefore, we focused on automatically generating appropriate content and process scaffolding based on sensors in the learning materials\(^2\).

The WoZ learning materials were replaced with sensorised versions that used simple embedded sensors and an overhead camera to automatically measure the tilt of the balance (left / right / centred) and the location of pots (which weight on which pin) and supporting blocks (present / removed). These simple sensors captured the state of the materials regarding these aspects with an accuracy of 96%. The system used a screen to display tasks and children entered their answers by pressing physical buttons. The Zeno robot was replaced with a simplified static robot embodiment (a plastic teddy bear) to encourage children to focus primarily on its scaffolding behaviours as opposed to its social behaviours and appearance. Based on sensor data and the child’s button presses, the RECAL system recognised which step of the

---

1DG 13: The robot could engage with the child in activities beyond learning
2Recommendation 1 - Designing educational behaviours: The robot should offer appropriate content support and process support
inquiry cycle children were in. It would then automatically generate appropriate scaffolding and feedback. The robot was explicitly presented to the child as a more knowledgeable other that would help them complete the assignment. Within that role it offered two forms of scaffolding. Firstly, it offered process scaffolding by strictly monitoring and enforcing the task structure, and by issuing reminders if children skipped steps or forgot task instructions. Secondly, it offered content scaffolding by providing simplified analogies to the children when they were unable to formulate hypotheses or conclusions. Additionally, it offered motivating, supportive, and friendly feedback while children progressed through the task.

The system was tested with children of 5-10 years old and we used semi-structured interviews to capture their impressions of the interaction. The study recorded logbooks of sensor values which were checked against observations to determine their accuracy. The design of the sensorised learning materials and corresponding scaffolding behaviour was presented as part of Davison et al. [48]. Detailed validation results are presented in Papenmeier [163]. Here we briefly discuss the relevant results from semi-structured interviews.

Regarding the educational level, children mentioned that the robot's support helped them judge the outcome of the tasks. They also mentioned that the robot helped them concentrate on the task and offered them useful instructions, explanations, and feedback. Regarding the relational level, children praised the robot's niceness and his motivational and social support. Furthermore, we conducted an exploratory content analysis of their descriptions of the robot. We speculate that the robot's positive feedback and encouragement gave children a sense of achievement and success while progressing through challenges.

5.6.4 Variant IV: A robot for longitudinal in-the-wild deployment

The work conducted with this variant has been submitted as Davison et al. [52, 51] and is discussed in Chapters 7 and 8 of this thesis.

Using the sensorised learning materials from Variant III we designed a learning intervention that could be deployed in an unsupervised manner in classrooms over extended periods of time. For this purpose, we focused on designing dialogues and behaviours that were robust enough for the robot to autonomously interact with the child without intervention or assistance from a researcher.

This variant was achieved by combining the Zeno robot's social repertoire from Variant I with the autonomous sensors and scaffolding from Variant III. The primary aim was to offer tailored praise, support, and encouragement to the child as they progressed through several assignments of increasing difficulty.

Most supplementary design choices were adopted from Variant I and were modified to work autonomously where necessary. Additionally, to signal that it could see and interpret what the child was doing, the robot would gaze reactively towards the child's actions as they manipulated the learning materials. Finally, we made the system more robust so that it could be used unsupervised, in a real classroom setting. On the one hand, we simplified somewhat the interaction repertoire of the robot. For

---

1DG 2: A robot should explicitly take on a certain role
2DG 4: A robot operating in the role of a more knowledgeable other could offer tutoring to the learner
3DG 11: The robot could offer emotional support
4DG 9: The robot should use appropriate signalling behaviours for the coordination of joint activities
instance, it no longer asked for verbal explanations, but instead asked children to select their hypotheses, observations, and conclusions from a multiple-choice list on a tablet interface. On the other hand, we added more variety to the exercises and gave children more freedom in how they could go through them. For instance, children could skip or repeat certain steps of the inquiry cycle as they preferred. In these cases the system reminded them about the proper order but did not necessarily prevent them from continuing.

This variant was deployed in classrooms during a four-month study with 6-10 year old children. We used a combination of semi-structured interviews, in-task self-reported measures, and automatically recorded usage statistics to follow the development of children’s interactions over time. We found that the robot’s praise had a positive effect on the children’s mindset in learning. The main results of the study are presented in Chapters 7 and 8.

5.7 Reflecting on the variants against the guidelines

Implementing and testing the four RECAL system variants gained us a first impression regarding the value of the guidelines for designing social behaviours for educational robots. Although we cannot proclaim that following these guidelines will always lead to intended or envisaged interactional effects, we found them a useful tool to structure and focus the design process. In each of the variants we addressed a specific subset of the design guidelines that were closely related to the social interactions that we aimed to elicit from participating children. In all cases the robot was successful: in Variant I the robot elicited meaningful verbal explanations, in Variant II it engaged children in a dancing activity, in Variant III children felt empowered by the robot’s scaffolding, and in Variant IV the robot successfully kept children engaged throughout a four month period while its praise and encouragement positively affected their mindset. Furthermore, throughout the various studies we observed that children described the robot as a social being of sorts and displayed a multitude of rich social behaviours towards it.

We suggest that designers continue to explore the applicability of the suggested guidelines in other educational interaction contexts, focusing on interactions that we have not yet investigated in the four variants outlined above. Firstly, the humanoid robot Zeno that was used as part of the platform was unable to physically manipulate learning materials; most of the collaborational interactions that we designed were therefore limited to gaze-shifts and speech. A robot with a different embodiment could be designed to make use of more intricate forms of collaborational behaviours. For example, a robot capable of manipulating and handing over physical objects (e.g. [91, 229]) may be designed to facilitate joint actions in an interdependent task and show prosocial actions towards collaborators.

Secondly, in several of the design guidelines we made suggestions related to the background story of the robot. Although we did not explore this aspect in detail, we found that children would occasionally make references to the background story of the robot or would use their imagination to describe such a story. For example, after playing with Variant II of the system (where children did an extracurricular dancing activity with the robot) they would imagine all sorts of other activities that the robot

---

1DG 6: To support a level of interdependence, the robot and child should facilitate each other’s actions

2DG 12: The robot could show prosocial behaviour
could do. We imagine that a robot could self-disclose its own background story as the interaction progresses (e.g. see the ALIZ-E project [58]) and may leverage this to disclose its own experiences and goals to the child\(^1\). In future work it could be interesting to investigate how a background story may influence children’s perceptions of the robot and how it could shape their collaboration efforts.

Finally, in the various CAL system variants we created dialogue models that would generate appropriate feedback and scaffolding. However, these models were not yet fully adaptive to individual children’s needs and preferences. During interviews we found indications that some children would have liked the robot’s help to be less repetitive and preachy. For example, by toning down the amount of instructions when a child performs well on a certain repetitive task. To address this, designers of educational robot interventions could offer more personalised scaffolding and interaction patterns [6, 131, 183].

5.8 Conclusion

In this chapter we described the research platform, consisting of sensorised learning tasks, a modular technical system architecture, and a collection of measures. We have briefly shown how this research platform was used to run studies with children and educational robots in a variety of settings, ranging from supervised single-sessions to autonomous in-the-wild longitudinal interactions. In the next chapters we present several of these studies in more detail.

\(^1\)DG 5: To express an intention for collaboration, the robot should disclose its own goals and should discuss task objectives, strategies, and joint actions
“There,” Chucky thinks to himself, “my very first learning task.” Though all his training has prepared him for this moment, he suddenly feels a bit scared. “I don’t know how this thing works! I’ve never even used a balance scale before, what if the kids ask me for help?”

But there is no time to second-guess, as Mees has already walked up and taken a seat in front of the task.

“Hi, I’m Chucky, wanna play?” The kid nods and reaches for one of the weighted pots. “So, have you done this before?” Chucky asks, “This is my first time, I might need your help to understand what is going on.”

“Oh-huh,” he nods, “I have done this one like a million times! It’s super easy, let me show you.” Mees starts reading the assignment and places various pots on the scale. “Ok, now make a prediction about the answer.”

Chucky instinctively thinks “Could it be forty-two?” but holds his tongue, “Nah, too obvious.” “What do you think?” he asks instead.

“Well, I think that side will go down,” Mees says while he removes the supports. “AH YES! See, I was right! That’s because that pot is much heavier!” He looks up at the robot and asks “Hey Chucky, what do you think?”

“Hmm, very interesting!”

6.1 Introduction

Results from the contextual analysis revealed that children typically engaged more often in learning behaviours of a physical nature than those of a cognitive nature. In this chapter we explore whether we can take advantage of a robot’s unique set of features to promote more cognitive behaviours: eliciting elaborate explanations from
the child. From the related work discussed in Chapter 2 we found that a social robot has valuable social characteristics that can enhance educational technology, enabling it to engage in rich social interactions. Robots and other social agents can quite naturally be given elaborate background stories and personas, they can develop and adapt their social abilities over time, can engage in interactive social (storytelling) behaviours, and can take on richer and more diverse roles in the learning process. Additionally, a robot is physically present in the same physical space in relation to the learning materials. This enables the robot to more naturally interact with the learner in the context of the task, for instance, by gesturing, gazing or pointing. Due to the robot's nature of being a physically embodied social agent, we expect it to be an appropriate tool to carry out certain social acts in learning.

Specifically, in the study presented in this chapter we were interested in a robot's influence on the child's explanatory behaviour in the context of a specific inquiry learning task. We explored the effect of extending a CAL system with a social robot on the explanatory behaviour of young children. We leveraged the inherent social nature of the robot to present it as an interactive co-learner with a background story. While guiding the inquiry learning task the children engaged in, the robot attempted to trigger verbal explanations from the child. Furthermore, the robot would provide feedback and ask questions about the inquiry task.

Explanatory behaviour is associated with gaining a greater understanding of one's own ideas and knowledge. For the past 15 years, research has proven that explaining leads to a deeper understanding when learning new things [39, 42, 75, 78]. There are two forms of explaining: (1) in solitary learning situations, where the learner explains the subject of interest to themself, which is called self-explaining, and (2) in collaborative learning situations, where a learner explains the learned subject to another person, which is called interactive explaining [168]. On the one hand, several studies have provided successful examples of self-explanation activities [39, 209]. Yet, providing self-explanations has one important disadvantage: it is not very intuitive to provide detailed explanations to oneself. On the other hand, Rittle-Johnson et al. [174] found that when children explained to another person (in this case their mother) who was only listening, it had a more positive effect on learning outcomes than self-explaining. A social partner can range from a partner who is just listening, to an interactive partner who provides support and feedback to the learner [78].

6.2 Research Question

The purpose of this study was to assess the value of a social robot for supporting a child’s verbalization of their knowledge and reasoning process during an inquiry learning task. As a baseline condition, we created a CAL system that guided a child through an exercise, where feedback and questions were read out loud by the system, delivered through a readily available PC speaker. In the experimental condition, the same feedback and questions were instead delivered through the social robot of the RECAL system, which was designed to leverage its embodied, social nature in the best possible way.

The main research question is: How do feedback and questions delivered through the RECAL system affect children’s explanatory behaviours?

We expect that a robot, with all its social expressive capabilities, will be more effective at eliciting verbal explanatory behaviour from children compared to the
Now we’re talking: Learning by explaining your reasoning to a social robot

Figure 6.1: The experiment setup illustrating two conditions: 1) Baseline no-robot CAL system, where feedback and questions are delivered through a wireless speaker; and 2) CAL system enhanced with a social robot, where identical feedback and questions are delivered by the robot through a wireless speaker attached to its back. In both conditions the CAL system was controlled by a dialogue engine running on a laptop, hidden out of view of the children.

‘less-social’ baseline CAL system. To measure the expected effect on explanatory behaviour, we focus on three dependent variables: 1) verbalisation duration; 2) number of aspects mention in the explanations (breadth); and 3) number of relations explicitly mentioned by the children in their explanations (depth). Based on the research on interactive explaining (e.g. [168]) we expect that children would provide more extensive and detailed explanations if they explained to the social robot, which would lead to a longer duration of verbalization. Furthermore, we expect that children would provide more thorough (i.e. broader and deeper) explanations to the robot.

6.3 Methodology

We conducted a study that employed a between-subjects design with two conditions as illustrated in Figure 6.1. In the baseline no-robot condition children performed an inquiry learning task without a robot, using only the basic CAL system. The system provided spoken assignments, feedback, and questions that were played through a wireless speaker placed on the table. A volunteer recorded the voice of the system beforehand to ensure that the verbal utterances were articulated clearly and understandably. We chose to include verbal support in the baseline system since not all children were able to read fluently and to ensure that the two conditions are as similar as possible (the robot also speaks). Children were instructed to provide their answers to the stated questions verbally.

In the robot condition children performed the same inquiry learning task, using the same baseline CAL system that was now extended with a social robot. In this condition the robot provided the same assignments, feedback, and questions as was done by the CAL system in the baseline condition. The same audio recordings were
played through the same wireless speaker that was now attached to the back of the robot. The behaviour design of the robot was informed by design guidelines emerging from an extensive contextual analysis of inquiry learning tasks with our target user group, reported in Chapter 4 [53]. While the child interacted with the robot and the learning task, the robot displayed the following behaviours: 1) **Facial expressions**: when children progressed through assignments the robot showed happy expressions, and directly after children performed the inquiry experiment the robot showed an amazed expression; 2) **Interactive gaze**: the robot gazed towards the child when speaking, the robot gazed towards the tablet when the child pressed a button or when a task appeared, and the robot gazed towards the learning materials when the objects were being manipulated; and 3) **Lifelike behaviours**: the robot blinked at random intervals, and lip-synchronization was added to give the impression that the robot was speaking. Some of the robot’s motors (in particular those in its limbs) are relatively noisy. Therefore, so as not to distract from the learning experience, we chose not to use full-body animations (such as pointing).

Furthermore, the design guidelines that followed from the contextual analysis suggested how the robot should be introduced to children [53]. The robot was introduced to the children as a peer but with well-developed inquiry skills (i.e. he knew how to perform the inquiry task, but did not know the correct answers yet). We presented the robot with a background story (similar to e.g. [126]), thereby contributing to its social characteristics. The background story of the robot was as follows: “It is a student from a planet far away. It is now on earth because it has an assignment from its teacher to study the effects of balance on earth. The robot wants to explore this phenomenon with like-minded people: children!”

### 6.3.1 Components of the CAL system

This study used system Variant I as described in Section 5.6. The experiment setup, as illustrated in Figure 6.1, consisted of the following components:

**Balance scale** The participants’ assignment was to explore the concept of balance (i.e. the moment of force) using a balance scale similar to the one used during the contextual analysis (see Figure 6.2). Children received a wooden balance with four pots: two red pots and two yellow pots. The colour of the pots represented their weight; red was heavier than yellow. Two wooden blocks could be placed under the balance to prevent the balance from tilting when placing or removing the pots. When children had finished placing the pots on the balance, they could remove the wooden blocks and observe whether the balance tilted to the left, to the right, or remained horizontal.

**Tablet** Children in both conditions used a Google Nexus 10 tablet, as shown in Figure 6.3. This tablet was used to display the assignments using pictures and text, and was used as input for the system. Children could indicate that they wanted extra help or continue to the next assignment by pressing one of two buttons on the tablet: a question mark for help and a green arrow to continue. When children pressed the question mark, the CAL system generated a more elaborate explanation depending on the current phase of the learning task.
Figure 6.2: The balance, coloured pots and wooden blocks used in the inquiry learning task.

Figure 6.3: The tablet interface used during both conditions. Translated text: “We are supposed to investigate what happens when we place two pots on the balance. Before we can begin with the experiment we have to prepare everything. Start by placing a yellow pot on pin 1 and a red pot on pin 5.”

Wireless speaker A small (approx. 10x10x3cm) wireless Philips BT2500B Bluetooth speaker was used as audio output for the CAL system. An identical speaker was used in both conditions, to ensure uniform audio quality. In the baseline no-robot condition, the speaker was placed on the table. In the robot condition, the speaker was attached to the back of the robot.

Social Robot Participants in the robot condition worked with the humanoid R25 Robokind robot ‘Zeno’ (see Figure 6.4). This robot is specifically suitable for simulating human facial expressions such as amazement and happiness. The identical audio recordings for the spoken assignments, feedback and questions, as used in the baseline CAL system, were now delivered by the robot. To rule out any influence of sound quality, the wireless speaker was attached to the back of the robot, replacing its built-in speaker. The robot displayed lip-synchronization movements in alignment with the
pre-recorded speech, making it appear as if the robot is speaking. As describe above, the robot displayed several behaviours during the interactions (facial expressions, interactive gaze, and life-like behaviours). The robot's behaviours were specified using Behaviour Markup Language (BML) [110] which was parsed and executed by AsapRealizer [172]. BML behaviours were generated at appropriate times during the interaction by the system's dialogue engine.

6.3.2 Inquiry learning assignments with the balance

The children received four consecutive assignments. These assignments started easy and became more difficult as children progressed through the tasks. In the first assignment, children used two pots of the same weight and placed them at the same distance from the pivot point on both sides, resulting in the balance staying horizontal. In the second assignment, children used two pots of the same weight, but the distance to the pivot point was varied. As a result, the balance tilted to one side. In the third assignment, the weight of the pots was varied but the distance to the pivot point was equal on both sides. The result was that the balance tilted to either one of the sides. In the final assignment, both the weight of the pots and the distance to the pivot point were varied. The result was that the balance remained horizontal.

In every assignment, children followed simplified steps of the inquiry cycle: prepare, predict, experiment, observe, and conclude. In the prepare phase children placed the pots on the balance, according to the assignment as was displayed on the tablet. In the predict phase children were asked by the system to verbally state their hypothesis about what would happen to the balance if the wooden blocks, that prevented the balance from tilting, were removed. In the experiment phase children removed the wooden blocks. In the observe phase children were asked by the system to tell what happened to the balance (tilted left, tilted right or remained horizontal). In the conclude phase the system asked the children to explain why they thought the balance was in this position.

6.3.3 Dialogue engine

In both conditions the CAL system was controlled by a dialogue engine running on a laptop, hidden out of view of the children. This central dialogue engine module was responsible for managing all aspects of the inquiry learning task: 1) monitoring the progression of the child in the task; 2) transitioning between the phases of the
Now we’re talking: Learning by explaining your reasoning to a social robot

Figure 6.5: Diagram of the states and transitions in the preparation phase of an inquiry learning task. Other phases of the task followed a similar structure of initial, help, error, and success states.

task; 3) generating images, buttons, and text to be displayed on the tablet; and 4) playing the audio files (pre-recorded speech) through the wireless speaker. The dialogue engine was constructed as a Finite State Machine (FSM) that modelled the interaction of the child with the learning materials as a collection of states, input events, transitions, and outputs.

6.3.3.1 States

Inquiry phases of the learning process of the child were represented as collections of states in the FSM. For example, at the start of each task, in the prepare phase, children had to place weighted pots on specific pins (e.g. a yellow pot on pin 1 and a red pot on pin 5) and wooden blocks under either side of the balance. The wooden blocks prevented the balance from tipping over and had to remain in place until the experiment phase. The FSM used four states to model the preparation phase, as shown in Figure 6.5: 1) the initial state containing the basic assignment instructions; 2) a help state for when the child requests or needs (i.e. when they remain idle) additional help; 3) an error state for when the wooden blocks are removed prematurely; and 4) a success state for when the pots are placed correctly.

6.3.3.2 Events and transitions

The FSM would transition between individual states based on certain input events. Transitions between states were triggered based on the following input events: i) the child manipulates the physical learning materials (i.e. placing / removing the weighted pots and placing / removing the wooden blocks); ii) the child presses buttons on the tablet (i.e. pressing the help button and pressing the continue button); iii) speech activity levels (specifically, state transitions were only triggered on an end of speech signal, not on speech content); and iv) a system timer expires (i.e. the child remains idle for a certain duration, between 10-60 seconds, depending on the phase of the
task). These transition triggers were chosen in this way because they do not require complex sensing by the system and are in principle feasible to detect automatically. Figure 6.5 shows the following events and state transitions: a) the system transitions to the success state when the pots are placed correctly according to the task; b) the system transitions to the help state when the child uses the wrong pots, places pots on incorrect locations, presses the help button, or when the system timer expires; c) the system transitions to an error state when the child prematurely removes the wooden blocks from under the balance; and d) the system transitions to the help state when the child corrects an error (i.e. by placing the blocks back under the balance).

6.3.3.3 System outputs

The system generated an output when transitioning into a state, as a response to the child’s actions. The system used the following outputs: 1) updating the interface of the tablet (i.e. showing / hiding image, text, and buttons); and 2) making a verbal utterance (i.e. reading the instructions aloud and offering feedback or encouragement). For example, during the prepare phase the system would offer the following outputs: 1) when transitioning to the initial state, the system would generate an assignment and would display an image, a help button, and a brief textual instruction on the tablet, which would be read out loud; 2) when transitioning to the help state, the system would display more elaborate textual instructions on the tablet and read them out loud to the child; 3) when transitioning to the error state, the system would offer a textual and verbal reminder to the child to keep the wooden blocks underneath the balance during the preparation phase; and finally 4) when transitioning to the success state, the system would display the continue button and would offer verbal encouragement.

In later prototypes of the system the dialogue engine was capable of automatically incorporating input from the child’s tablet (i.e. button presses), external sensors (e.g. to measure the state of the learning materials [48]), and perception modules (e.g. voice activity detection [64]). Based on these inputs, the dialogue engine enabled the CAL system to follow the child’s actions in the learning task and respond appropriately. However, in the prototype reported in this chapter we used a Wizard of Oz approach to simulate several sensors and perception modules.

6.3.4 Wizard of Oz

We opted for a Wizard of Oz (WoZ) approach to provide input to the system regarding the child’s actions in the task. Using this approach a researcher simulated certain sensor values according to a predefined protocol. The simulated sensor values were input into the CAL system’s dialogue engine, which then automatically triggered any applicable state transitions and system responses based on its FSM models.

The WoZ protocol was constructed during early pilot tests and was intended to cover the core interaction and common edge cases [219]. The full WoZ protocol for both conditions is included in the next section. The protocol was identical for the baseline no-robot condition and the robot condition. Table 6.1 shows a representative transcript of a children working on a task from both conditions. During the pilot tests the robustness of the protocol was validated by having a non-Dutch speaking researcher (who did not understand the content of the children’s speech) successfully wizard sessions of Dutch children. As part of the core interaction the wizard would simulate the following sensor data:
1. Have the weighted pots been removed or are they placed on correct / incorrect locations according to the task?

2. Are the wooden blocks removed or are they placed underneath the balance?

3. In the hypothesize, observe, and conclude phases: has the child stopped speaking for 2 seconds?

Additionally, the dialogue engine would keep track of the time that children remained idle in the task (meaning that they did not speak, press buttons, or manipulate the learning materials). After a fixed timer duration had elapsed, the “trigger a timer event” button on the WoZ control interface would light up. Following the protocol, the wizard would press this button only if the child was actually idle, but would ignore the button if the child was actively working on the task\(^1\).

Furthermore, in rare cases the child would turn towards the researcher for assistance or would very explicitly ask the system (verbally) what they had to do. In these cases, following the protocol, the wizard would manually trigger a transition to the help state by pressing the “offer additional help” button\(^2\).

Finally, in the conclude phase of the task, the system would ask the child “Can you explain why the balance is like this?” Most of the time the child would offer an explanation and the wizard would indicate when they stopped talking. Regardless of the actual content of their explanation, the system’s default response was “That’s very interesting!” However, on occasion children would simply answer either “No” or “I don’t know”. In these two situations the system’s default response would be inappropriate. Therefore, according to the protocol, the wizard could select the option: “conclusion unknown” enabling the system to respond with “OK”, after which the task would continue as usual\(^3\). This was the only situation where the wizard had to interpret the content of the child’s speech.

### 6.3.5 WoZ Protocol for both conditions

**Preparation phase**

Initiate state [system speech]: “You can start by placing a red pot on pin 2 and a red pot on pin 5.”

**Response options:**

1. Child removes blocks → Wizard presses button: “Blocks removed” → System goes to error state

2. Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state.

3. Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state

---

\(^1\)The trigger a timer event button was pressed a total of 10 times by the WoZ during the study.

\(^2\)The offer additional help button was pressed a total of 1 time by the WoZ during the study.

\(^3\)The conclusion unknown button was pressed a total of 36 times by the WoZ during the study.
Table 6.1: Transcripts of sessions in the baseline no-robot condition and the robot condition. The dialogue engine’s FSM models, the system’s utterances, and the WoZ protocols were identical in both conditions.

<table>
<thead>
<tr>
<th>Baseline No-Robot condition</th>
<th>Robot condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prepare phase</strong></td>
<td></td>
</tr>
<tr>
<td>System: “Let’s start by placing a red pot on pin 2 and a red pot on pin 6”</td>
<td>Robot: “Let’s start by placing a red pot on pin 2 and a red pot on pin 6”</td>
</tr>
<tr>
<td>Child: Places pots</td>
<td>Child: Places pots</td>
</tr>
<tr>
<td>WoZ: Presses button [pots are placed correct]</td>
<td>WoZ: Presses button [pots are placed correct]</td>
</tr>
<tr>
<td>System: “Great!”</td>
<td>Robot: “Great!”</td>
</tr>
<tr>
<td><strong>Predict phase</strong></td>
<td></td>
</tr>
<tr>
<td>System: “Later, you may remove the blocks. First, can you explain what you think will happen when you remove the blocks?”</td>
<td>Robot: “Later, you may remove the blocks. First, can you explain what you think will happen when you remove the blocks?”</td>
</tr>
<tr>
<td>Child: “I think the red pot on pin 6 will go down”</td>
<td>Child: “I think that... the right side will go down a bit, because on the right side it [the pot] is much further”</td>
</tr>
<tr>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [hypothesis given]</td>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [hypothesis given]</td>
</tr>
<tr>
<td>System: “OK”</td>
<td>Robot: “OK”</td>
</tr>
<tr>
<td><strong>Experiment phase</strong></td>
<td></td>
</tr>
<tr>
<td>System: “Now you're going to do the experiment. You may remove the two blocks from under the balance.”</td>
<td>Robot: “Now you're going to do the experiment. You may remove the two blocks from under the balance.”</td>
</tr>
<tr>
<td>Child: Removes wooden blocks</td>
<td>Child: Removes wooden blocks</td>
</tr>
<tr>
<td>WoZ: Presses button [blocks are removed]</td>
<td>WoZ: Presses button [blocks are removed]</td>
</tr>
<tr>
<td>System: “Great!”</td>
<td>Robot: “Great!”</td>
</tr>
<tr>
<td><strong>Observe phase</strong></td>
<td></td>
</tr>
<tr>
<td>Child: “6 went down and 2 went up”</td>
<td>Child: “It went to the right because it is further away”</td>
</tr>
<tr>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [observation given]</td>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [observation given]</td>
</tr>
<tr>
<td>System: “That’s interesting!”</td>
<td>Robot: “That’s interesting!”</td>
</tr>
<tr>
<td><strong>Conclude phase</strong></td>
<td></td>
</tr>
<tr>
<td>System: “Can you explain why the balance is like this?”</td>
<td>Robot: “Can you explain why the balance is like this?”</td>
</tr>
<tr>
<td>Child: “Because... ehhm well first they were on the same pins and now one [pot] is on the other and the balance goes... ehh... goes a bit down and then the pot, I think, goes down automatically”</td>
<td>Child: “Because at two it is less far from the thing [midpoint] for example... and at 4... the pot that is on pin 6 is much further away”</td>
</tr>
<tr>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [conclusion given]</td>
<td>WoZ: Waits for child to finish speaking for two seconds, then presses button [conclusion given]</td>
</tr>
<tr>
<td>System: “That is very interesting!”</td>
<td>Robot: “That is very interesting!”</td>
</tr>
</tbody>
</table>
4. Child places pots incorrectly → Wizard presses button: “Pots placed incorrect” → System goes to initial state (repeats assignment).

5. Child places pots correctly → Wizard presses button: “Pots placed correct” → System goes to success state (green continue button appears on the child's tablet).

Predict phase

Initiate state [system speech]: “Later you may remove the wooden blocks. First, can you explain what you think will happen when you remove the blocks?”

Response options:
1. Child removes blocks → Wizard presses button: “Blocks removed” → System goes to error state
2. Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state.
3. Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state
4. Child places pots incorrectly → Wizard presses button: “Pots placed incorrect” → System goes to initial state (repeats assignment).
5. Child provides a response (content irrelevant) → After 2 seconds Wizard presses button: “hypothesis given” → System goes to success state (green continue button appears on the child’s tablet).

Experiment phase

Initiate state [system speech]: “Now you are going to do the experiment. You may remove the two block from under the balance”

Response options:
1. Child removes blocks → Wizard presses button: “Blocks removed” → System goes to success state (green continue button appears on the child’s tablet).
2. Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state.
3. Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state

Observe phase

Initiate state [system speech]: “Now you are going to do the experiment. You may remove the two block from under the balance”

Response options:
1. Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state.

2. Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state

3. Child provides response (content irrelevant) → After 2 seconds Wizard presses button “Observation given” → System goes to success state (green continue button appears on the child’s tablet).

Conclude phase

Initiate state [system speech]: “Can you explain why the balance is like this?”

Response options:

1. Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state.

2. Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state

3. Child provides negative response (No or I don’t know) → After 2 seconds Wizard presses button “Conclusion unknown” → System goes to success state (green continue button appears on the child’s tablet).

4. Child provides positive response (gives conclusion) → After 2 seconds Wizard presses button “Conclusion given” → System goes to success state (green continue button appears on the child’s tablet).

6.3.6 Exit Interview

After children had completed the experiment, we conducted semi-structured exit interviews. These provided us with an indication whether children perceived the robot as more social compared to the baseline CAL system. In the baseline no-robot condition, from the children’s perspective, the CAL system was represented by the tablet. Therefore, during the interviews in the no-robot condition, we referred to the tablet when asking about their experience. The following questions were used in both conditions:

1. *Did you enjoy it?* This was asked as a warm-up question for the children to become accustomed to the interview process.

2. *What did you like most?* This was the second warm-up question. This question was asked to trigger children to recall the activity. We expected that the most prevalent event from their experience would be mentioned here.

3. *Can you tell me something about the robot/tablet?* This question was asked to see what children found the most prevalent about the robot/tablet.
4. When you go home after school, what will you tell your parents/siblings about what you did here? Again this question was asked to see which experience was the most prevalent according to the children and worth telling other people about.

5. Could the robot/tablet see what you were doing? This question was asked to see if children thought they were visible to the technology.

6. Could the robot/tablet hear you? This question was asked to see if children thought they were heard by the technology.

In the robot condition the following additional questions were asked specifically about the robot:

7. Did the robot help you or did you help the robot? This question was asked to see how children related to the robot in context of the task.

8. Do you think the robot is smart? This question was asked to see how children estimated the cognitive capabilities of the robot.

9. Do you think the robot is friendly? This question was asked to see how children regarded the robot on a social level.

10. Do you think the robot thinks you are friendly? This question was asked to see how children regarded the robot on a social level.

11. How old do you think the robot is? This question was asked to see whether children would give him a “human” or “device” age. This could indicate how they perceived the robot.

Additionally, children in the baseline no-robot condition were given the opportunity to also do one additional assignment with the robot, after fully completing the experiment and interview. These children were then asked to compare their experience with the robot and the tablet:

7. What do you think is better? With the tablet or with the robot? This question was asked to see which device the children preferred in the context of this task.

6.3.7 Picture task

A picture task\(^1\) was used as a second measure to gain insight into how the CAL system and the robot were perceived by the children. A variety of pictures were selected to cover a broad spectrum of properties, as shown in Figure 6.6. They included entities that spanned the space of non-social (e.g. car, notebook, laptop) to less social (e.g. a teddy bear, dog) to more social (e.g. friends, teacher), and inanimate (e.g. teddy bear, note book, laptop) to animate (e.g. dog, cat). Furthermore, the pictures of a teacher and friends could represent a variety of dualities commonly found in educational settings, like social roles, authority/status, and knowledge levels. Finally, objects were included that children may associate with everyday life (e.g. car) or with school, being either technical (e.g. laptop) or non-technical (e.g. note book) in nature.

The pictures were used primarily as a way to advance the conversation and focus on particular aspects by affording children to draw comparisons using terms they may be

\(^1\)The picture task reported here was the first exploratory implementation of this tool. A more optimised and validated iteration of the measurement instrument is reported in Charisi et al. [35].
familiar with. Children were asked to select a picture from a given set that they thought did or did not fit to the tablet \(^1\) or the robot. Note that there were not necessarily correct or incorrect answers, as we were mostly interested in the explanations children gave as to why they selected certain pictures. We also emphasised this point when interviewing the children.

When starting the picture selection task the set of images were presented on the table. Children first received a picture of either the tablet or the robot. They were then asked to select a picture that, according to them, best fitted with the tablet/robot, and explain why they chose that specific picture. Their explanation was a key element for further analysis. This process was repeated for a second picture. Then children were asked to pick a picture that, according to them, did not fit with the tablet/robot and explain why they chose that picture. This was also repeated for a second picture. The picture task was embedded in the exit interview and was introduced after children had answered question six. When children finished the picture task the experimenter continued the interview with question seven.

The exit interviews, including the picture task, were video-recorded. The answers children gave were first transcribed. Then, the data was structured according to categories that emerged from the data following an inductive grounded theory approach. Analysis of the categories gave insights in the responses children gave per question in the two conditions and an overview of the responses children gave during the picture task per condition. Three researchers coded these answers. An overview of the responses is presented in section 6.5.4.

### 6.3.8 Procedure

Prior to the experiment, ethical approval was granted by the ethical boards of the Behavioural, Management and Social sciences (BMS) and the Electrical Engineering, Mathematics and Computer Science (EEMCS) faculties of the University of Twente. Permission forms were distributed to the children’s parents through the school’s regular communication channels. Sessions took place in a separate room in the school during regular school hours. The children first received an introduction by the experimenter, which took approximately 5 minutes. In the robot condition, the experimenter first introduced the Zeno robot and its background story. In both conditions the experimenter told the children that they should provide their answers by talking out loud. The experimenter then showed how the pots were placed on the balance and explained the function of the two buttons on the tablet (the question mark and the green arrow). After this introduction the experimenter walked away and sat somewhere out of sight in the same room. The children then started with the first assignment. During the experiment, the experimenter served as Wizard of Oz controller.

---

\(^1\)In the no-robot condition where children worked with only the CAL system, from the children’s perspective, the entire system was represented by the tablet. Therefore, during the picture task we referred to the tablet when asking them to make a comparison.
Now we’re talking: Learning by explaining your reasoning to a social robot

Table 6.2: Participant demographics per condition.

<table>
<thead>
<tr>
<th></th>
<th>Age Mean (SD)</th>
<th>Nr. of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No-Robot</strong></td>
<td>8.2 (1.5)</td>
<td>22 (15/7)</td>
</tr>
<tr>
<td><strong>Robot</strong></td>
<td>8.1 (1.3)</td>
<td>24 (12/12)</td>
</tr>
</tbody>
</table>

It took children approximately 20 minutes to complete the assignments. Then the experimenter started the exit interview, including the picture task. Children from the baseline no-robot condition needed approximately 5 minutes to complete the interview and picture task. Children from the robot condition needed approximately 10 minutes.

In order to clarify how sessions went in practice we added two videos as supplementary materials to Wijnen et al. [220]. In the first video we highlight the setup and procedure of the experiment and the function of the Wizard of Oz. In the second video an actor re-enacts a transcript of a representative session from each condition, to illustrate how real sessions with children went in practice.

6.3.9 Participants

In total, 53 Dutch elementary school children participated in the study. Sessions which were incomplete or that contained missing data were removed, resulting in 46 sessions suitable for further analysis. The demographics for each condition are given in Table 6.2. The children were selected from two different schools in the same city and were randomly assigned to either the baseline no-robot condition ($N = 22$), or the robot condition ($N = 24$).

6.4 Annotating and analysing children’s speech

Every session was video-recorded from two angles to capture the full interaction scene. In addition, the dialogue system automatically collected log data that contained all input from the Wizard of Oz tablet and from the tablet used by the children, and every output action from the CAL system. The log data was used to automatically annotate the videos, using the language archive program ELAN [194]. The videos were then further manually annotated on two levels.

The first level was child speech (annotated manually for inquiry phases predict, observe and conclude) and contained two labels: verbalization and fillers. Verbalization was used when children provided verbal explanations and was used directly to assess the duration of verbalization. Fillers was used when children voiced utterances such as “ehmm...” that indicated that they wanted to say something, Fillers were only annotated when they occurred at the start of an utterance, preparatory to the child providing their actual explanation. We separated preparatory fillers from verbalization because we were primarily interested in measuring the length of the actual explanation.

The second level was system speech and contained three labels: giving explanation, asking question, and waiting for response. System speech was annotated automatically based on generated log files from the CAL system. Giving explanation was used when the CAL system gave an explanation or a verbal response to the child. Asking question was used when the system stated a question and was also annotated automatically. Waiting for response was used when the system had stated a question and was waiting
for a response from the child. The annotations were converted to numerical data and analysed using the R statistical analysis program\(^1\).

To gain a better understanding of the quality and thoroughness of children’s explanations, we coded the content of their explanations on two levels: 1) breadth—how many different things do they talk about; and 2) depth—how often they relate these things together.

6.4.1 Explanation breadth

To analyse the breadth of each explanation we investigated the number of facets that were mentioned. Using a grounded theory approach, we derived a coding scheme consisting of nine codes (the facets) through axial coding:

**Balance position** Used when a child explicitly made a remark about the current position of the balance. For example, “this side went down” or “it is in balance”.

**Weight of pots** Used when a child explicitly mentioned the (relative) weight of the pots. For example, “this pot is just as heavy as that one” or “there is more in this pot than in that one”.

**Distance of pots** Used when the child explicitly mentioned the (relative) distance of the pots to the pivot point. For example “their distance is the same” or “this one is all the way to the end”.

**Position of pots** Used when a child explicitly mentioned the (absolute) position of the pots on the balance. Note: this is different from mentioning the distance from the pivot point. For example: “this pot is on pin 1 and this one is on pin 4” or “the red one is here and the yellow one is there”.

**Example/counter example** Used when the child gave a (counter)example to clarify his/her explanation. For example, “the balance has now tilted to the right, but if this pot were here it would remain in balance”.

**Naive weight** Used when the child explicitly mentioned that one side of the balance was heavier, causing it to fall down. In this case the child did not necessarily show understanding of all principles of the balance work, but did show a naive understanding of the moment of force. For example: “the pots have equal weight but this side of the balance is heavier” or “this pot is heavier, but there is the same weight on both sides”.

**Circular reasoning** Used when a child explained the position of the balance without mentioning one of the variables (distance from the pivot point or weight of the pots). In other words, they were stating the obvious, without showing understanding of the underlying principles. For example, “the left side goes down because it is heavier” or “it is balanced because it weighs the same on both sides”. As opposed to the “naive weight” code, a child using circular reasoning does not explicitly mention the weight or position of the pots on the balance in their explanation.

**Other** Used when the child talked about something that could not be coded with one of the codes described above.

\(^1\)R Project for Statistical Computing: [http://www.r-project.org](http://www.r-project.org)
No answer  Used when a child did not know the answer or gave no explanation at all.

These nine codes covered all facets relevant to explaining the outcomes of the balance task. During content analysis, we coded whether a child talked about a specific facet during their explanation, not how long or what exactly they said about it. Consequently, each explanation was coded with respect to the facets that were mentioned in the explanation. In cases where children mentioned multiple facets in their explanation, this resulted in multiple codes. If a child repeated his/herself during an explanation, multiple identical codes were used. For example, an explanation such as “The red pot is heavier than the yellow pot. So then the balance tipped to the right, because the red pot is heavier.” would be coded with 2x Weight of pots and 1x Balance position.

Following the above approach, researcher A annotated all data and researcher B independently annotated 20% of the data. A Cohen’s K of 0.87 showed an excellent inter-rater agreement. Some confusion between annotators was found for codes distance of pots and position of pots. Closer examination of the coded fragments revealed an ambiguity in the wording of the coding scheme, which was consistently misinterpreted by researcher B. Correction of this ambiguity resulted in an updated Cohen’s K of 0.92, which indicates that the above collection of codes are a viable method for reliably coding and interpreting the content of children’s explanations.

6.4.2 Explanation depth

To analyse the depth of each explanation we coded the number of logical associations that were explicitly mentioned by the child connecting those facets. Explanations that incorporate associations (or inferences) seem to enhance students’ knowledge [168] and are important characteristics of scientific explaining [118]. Explicitly mentioning associations between facets may illustrate an understanding of the presence (or absence) of a relationship that may exist between the facets.

We annotated the use of logical deductive phrases to code whether an explicit association was drawn between facets of an explanation. Specifically, phrases that are causal, oppositional, or conditional in nature. Examples of such phrases are: “but”, “although”, “because”, “that’s why”, “and so”, “therefore”, “that means that”, “provided that”, “otherwise”, etcetera. On the one hand, an example of a shallow explanation could be: “The balance is horizontal. The red pot is heavy and is placed on pin four. The yellow pot is light and is placed on pin one.” Although this example covers three relevant facets, namely [Balance position], [Weight of pots], and [Position of pots], it does not explicitly illustrate a deeper level of understanding with regard to the logical relations that exist between the facets. On the other hand, a much deeper explanation using the same three facets could be: “Although the red pot is heavy, it is placed on pin four. The yellow pot is light and is therefore placed on pin one. That’s why the balance is horizontal.” This phrase explicitly draws logical associations between the mentioned facets with the keywords “although”, “therefore”, and “that’s why”. For this example we would annotate three associations.
6.5 Results

6.5.1 Duration of verbalization

Analysis for normality showed that the data for duration significantly deviated from a normal distribution in both conditions (Shapiro-Wilk, $p < 0.001$). Therefore, Mann Whitney U tests were used to determine the differences in duration between the two conditions.

The response duration of the child was measured as the duration (in seconds) of their responses to questions stated by the system (minus any leading fillers like “Uuhhm”). The Mann Whitney analysis showed that, overall, children verbalized for a significantly longer duration in the robot condition ($M = 5.7, SD = 2.5$) than in the baseline no-robot condition ($M = 4.2, SD = 2.4$), $U = 172.00, z = -2.019, p = .043$ (two-tailed).

In addition to the analysis of the overall results, we looked more in depth into the differences per inquiry phase (with Bonferonni correction: $\alpha = .0167$). The Mann Whitney U test showed no significant differences in the predict phase for duration (robot: $M = 4.6, SD = 2.6$, no-robot: $M = 3.7, SD = 2.2$): $U = 194.00, z = -1.525, p = .127$ (two-tailed). In the observe phase, the Mann Whitney U test showed no significant difference between the baseline no-robot and the robot condition for the duration of the verbalization (robot: $M = 5.5, SD = 3.4$, no-robot: $M = 3.4, SD = 1.8$), $U = 161.00, z = -2.270, p = .023$ (two-tailed). In the conclude phase there was no significant difference between the conditions for duration (robot: $M = 7.1, SD = 3.7$, no-robot: $M = 5.5, SD = 4.1$): $U = 186.00, z = -1.704, p = .088$ (two-tailed).

These results show that children verbalized for a longer duration to the robot. From this we conclude that in this study questions asked by a social robot yielded longer verbalizations than those asked by a baseline CAL system.

6.5.2 Duration of response versus explanation breadth

We carried out a content analysis of the children’s verbalizations by annotating the various facets that the children addressed in their responses. A preliminary analysis of children’s responses revealed that, in both conditions, the most elaborate explanations were given in the observe phase and conclude phase. We therefore focused annotating
Table 6.3: Explanation breadth: Number of times each code was used during annotation in each phase in the two conditions. For each code we report the mean and its standard deviation of the number of times that code was used during the annotation of each child.

<table>
<thead>
<tr>
<th>Code</th>
<th>Total per child</th>
<th>Mean (SD)</th>
<th>Total per child</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observe phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balance position</td>
<td>78</td>
<td>3.55 (0.96)</td>
<td>83</td>
<td>3.46 (1.02)</td>
</tr>
<tr>
<td>Weight of pots</td>
<td>7</td>
<td>0.32 (0.57)</td>
<td>22</td>
<td>0.92 (1.06)</td>
</tr>
<tr>
<td>Distance of pots</td>
<td>1</td>
<td>0.05 (0.21)</td>
<td>13</td>
<td>0.54 (0.72)</td>
</tr>
<tr>
<td>Position of pots</td>
<td>3</td>
<td>0.14 (0.47)</td>
<td>11</td>
<td>0.46 (0.72)</td>
</tr>
<tr>
<td>Example/counterexample</td>
<td>1</td>
<td>0.05 (0.21)</td>
<td>5</td>
<td>0.21 (0.72)</td>
</tr>
<tr>
<td>Naive weight</td>
<td>1</td>
<td>0.05 (0.21)</td>
<td>2</td>
<td>0.08 (0.28)</td>
</tr>
<tr>
<td>Circular reasoning</td>
<td>6</td>
<td>0.27 (0.88)</td>
<td>2</td>
<td>0.08 (0.28)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0 (0)</td>
<td>1</td>
<td>0.04 (0.2)</td>
</tr>
<tr>
<td>No answer</td>
<td>1</td>
<td>0.05 (0.21)</td>
<td>2</td>
<td>0.08 (0.41)</td>
</tr>
<tr>
<td><strong>Sub-total</strong></td>
<td>98</td>
<td>4.45 (0.8)</td>
<td>141</td>
<td>5.88 (2.31)</td>
</tr>
<tr>
<td><strong>Conclude phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balance position</td>
<td>8</td>
<td>0.36 (0.66)</td>
<td>10</td>
<td>0.42 (0.5)</td>
</tr>
<tr>
<td>Weight of pots</td>
<td>35</td>
<td>1.59 (1.18)</td>
<td>39</td>
<td>1.62 (0.97)</td>
</tr>
<tr>
<td>Distance of pots</td>
<td>14</td>
<td>0.64 (0.9)</td>
<td>31</td>
<td>1.29 (1.16)</td>
</tr>
<tr>
<td>Position of pots</td>
<td>19</td>
<td>0.86 (1.08)</td>
<td>20</td>
<td>0.83 (1.01)</td>
</tr>
<tr>
<td>Example/counterexample</td>
<td>2</td>
<td>0.09 (0.29)</td>
<td>7</td>
<td>0.29 (0.86)</td>
</tr>
<tr>
<td>Naive weight</td>
<td>7</td>
<td>0.32 (0.57)</td>
<td>6</td>
<td>0.25 (0.53)</td>
</tr>
<tr>
<td>Circular reasoning</td>
<td>6</td>
<td>0.27 (0.63)</td>
<td>7</td>
<td>0.29 (0.75)</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>0.18 (0.39)</td>
<td>6</td>
<td>0.25 (0.44)</td>
</tr>
<tr>
<td>No answer</td>
<td>19</td>
<td>0.86 (1.49)</td>
<td>15</td>
<td>0.62 (1.06)</td>
</tr>
<tr>
<td><strong>Sub-total</strong></td>
<td>114</td>
<td>5.18 (2.11)</td>
<td>141</td>
<td>5.88 (1.75)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>212</td>
<td>9.64 (2.63)</td>
<td>282</td>
<td>11.75 (3.35)</td>
</tr>
</tbody>
</table>

these two phases. In the predict phase children primarily stated their hypotheses, often without providing additional justification or explanation.

The results of the content analysis are shown in Figure 6.7 and summarized in Table 6.3. A Mann Whitney U test was used to investigate whether the differences between the number of codes in each condition was significantly different. First, for each child the aggregated number of codes was calculated over the whole session consisting of four assignments. We found that, overall, in the robot condition ($M = 11.75$, $SD = 3.35$) children mentioned more facets in their explanations than in the baseline no-robot condition ($M = 9.64$, $SD = 2.63$), $U = 143.50$, $z = -2.675$, $p = .007$ (two-tailed). In the observe phase we found a significant difference (with Bonferroni correction: $\alpha = .025$) between the no-robot condition ($M = 4.45$, $SD = 0.8$) and the robot condition ($M = 5.88$, $SD = 2.31$), $U = 157.00$, $z = -2.60$, $p = .009$. We found no overall significant difference in the conclude phase, $U = 186.00$, $z = -1.758$, $p = .079$. When we explore Table 6.3 we see that the facets balance position, weight of pots, distance of pots and position of pots are the most often mentioned facets in both conditions. This is interesting because these facets are crucial to mention if one
Table 6.4: Examples of explanations given in both conditions in the conclude phase.

<table>
<thead>
<tr>
<th>Baseline No-Robot condition</th>
<th>Robot condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclude phase</td>
<td></td>
</tr>
<tr>
<td>System: &quot;Can you explain why the balance is like this?&quot;</td>
<td>Robot: &quot;Can you explain why the balance is like this?&quot;</td>
</tr>
<tr>
<td>Child: &quot;I think the red [pot] is heavier&quot;</td>
<td>Child: &quot;Because the red pot is the heaviest and the yellow pots are the lightest&quot;</td>
</tr>
<tr>
<td>Child: &quot;Because one [pot] is on the rear and one [pot] is in the middle&quot;</td>
<td>Child: &quot;Because the pot that is on pin 6 is farthest away from the beam and on the blue side it [the pot] is on number 2&quot;</td>
</tr>
<tr>
<td>System: &quot;That is very interesting!&quot;</td>
<td>Robot: &quot;That is very interesting!&quot;</td>
</tr>
</tbody>
</table>

Table 6.5: Explanation depth: Number of logical association between facets in each phase per condition.

<table>
<thead>
<tr>
<th></th>
<th>No-Robot</th>
<th>Robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Observe phase</td>
<td>6</td>
<td>0.07 (0.14)</td>
</tr>
<tr>
<td>Conclude phase</td>
<td>73</td>
<td>0.84 (0.63)</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>0.45 (0.35)</td>
</tr>
</tbody>
</table>

wishes to fully explain the function of the balance.

These results indicate, firstly, that when children verbalized longer they mentioned more relevant facets in their explanations. Secondly, these results show that, overall, children touched upon more relevant facets when explaining to the robot than to the baseline no-robot CAL system. From this we conclude that when questions were delivered by a social robot, as opposed to a baseline CAL system, children were prone to give more detailed and relevant explanations.

6.5.3 Duration of response versus explanation depth

To investigate the depth of the children’s explanations we coded how often children drew logical associations between facets through the use of causal, oppositional or conditional deductive phrases. Table 6.4 provides some examples of explanations given by children in the conclude phase for both conditions.

We annotated a total of 140 logical associations in the robot condition and a total of 79 in the baseline no-robot condition. Table 6.5 lists the details per condition and per phase. A Mann-Whitney U test showed a significant overall difference between conditions, \( U = 156.00, z = -2.364, p = .018 \) (two-tailed). For explanations given in the observe phase we found a significant difference between the robot condition and the base-
Now we’re talking: Learning by explaining your reasoning to a social robot

line no-robot condition, $U = 141.00$, $z = -3.026$, $p = .002$ (two-tailed) (with Bonferroni correction: $\alpha = .025$). No significant difference was found in the conclude phase.

This indicates that children who interact with the robot make more (explicit) logical associations between relevant facets compared to children who worked with the no-robot baseline condition.

6.5.4 Exit interviews: was the robot perceived as social?

In the baseline no-robot condition, there were three sessions for which the log file data was not properly stored, and therefore excluded in the analysis of the verbalization content and duration. Otherwise, these sessions were completed and were therefore included in the exit interview and picture task analysis resulting in the analysis of responses from 25 children in the baseline no-robot condition and 24 children in the robot condition.

The first two questions in the interview were warming-up questions, and were asked primarily so the child could get used to the interview process. Responses to these two questions are reported as such and will not be further interpreted. To the first question: Did you enjoy it? Most children (44 of 49) answered that they enjoyed the experiment. To the second question: What did you like most? Children in both conditions reported that they enjoyed some part of the inquiry, such as: “I enjoyed placing the pots on the balance”. Some children also mentioned the interaction with the tablet/robot. For example “I can just talk to the robot” or “that thing (tablet) talked to me, which was funny”.

The third question: Can you tell me something about the robot/tablet? was asked to see what children found most noticeable about the robot or tablet. Most children in the robot condition gave a response regarding the interactive behaviour of the robot (10 of 24) for example “He [the robot] can explain things to me”. Furthermore some children mentioned the embodiment of the robot (8 of 24). Thereby focusing on two things: that the robot moved, and the size of the robot. Interestingly, children did not mention the facial expressions that the robot displayed. Additionally, children in the robot condition mentioned the role of the robot in the task (6 of 24). For example: “The robot taught me a lot”. In the baseline no-robot condition, children mostly talked about the interaction with the tablet (10 of 25). For example, “it was interesting that it could talk”. Additionally, children in the baseline condition mentioned the embodiment/design (5 of 25) of the tablet. For example, “There are some pictures here for if you don’t quite understand”.

To the fourth question: When you go home after school, what will you tell your parents/siblings about what you did here? The responses of the children in both conditions were mostly related to the description of the task (baseline: 11 of 25, robot: 15 of 24). For example, “that I have worked with the balance”. In the robot condition, the children also talked about the embodiment of the robot (7 of 24). For example, “I would tell that the robot could move and talk” or “that he is small and has a pleasant voice”. In addition, again several children in the robot condition mentioned the interactive behaviour of the robot (5 of 24). For example, “He [the robot] can talk”. In the tablet condition, only one child mentioned the embodiment of the tablet. Children were more focused on the interaction with the tablet (8 of 25). For example, “He [the tablet] talked to me and I didn’t have to type, but I could just talk to it”.

To the fifth question: Could the robot/tablet see what you were doing? Most children in the robot condition answered that they thought the robot could see them (19 of 24).
In the no-robot condition about half of the children said that the tablet could see them (12 of 25). We also asked why children thought the robot/tablet could (or could not) see them? In both conditions responses were related either to some technical aspect, such as, “I think he (the robot) has a camera in his eyes” or the reactive behaviour of either the tablet or robot, such as: “if I put the blocks under the balance, the green arrow appeared”. To the sixth question: Could the robot/tablet hear you? Most children in the robot condition thought that the robot was able to hear them (19 of 24). In the no-robot condition 18 of 25 children thought that the tablet could hear them. When asked to explain their answers, most explanations were related to the responsive behaviour of the system. However, occasionally children in both conditions also mentioned that it was some external factor that caused the robot/tablet to give a response, such as: “you heard me and controlled the robot”.

In the robot condition some additional questions were asked. To the question: Did the robot help you or did you help the robot? The majority of children responded that he/she and the robot helped each other. To the question: Do you think the robot is smart? Most children responded that they thought the robot was smart. To the question: Do you like the robot? Almost all children answered: “yes”. To the question: Do you think the robot likes you? Almost all children answered: “yes”. To the last question: How old do you think the robot is? 20 children gave the robot a “human age” between 5 and 12 years old, which is around the same age as most of the children. There were only 4 children who asked for clarification or responded in a way that indicated they were considering giving the robot a “device age”. For example, “I think he is not so old. We didn’t have robots yet in the year 2000”.

In the baseline no-robot condition, children had the opportunity to do one assignment with the robot after they finished the experiment. When children chose to do this, we asked them: What do you think is better? With the tablet or with the robot? 18 of the 22 children indicated that they preferred the robot. In most cases the children mentioned the appearance of the robot. For example, “he can move and he looks at you and he really says something!” There was only one child who said that they preferred working with only the tablet, although he/she did not explain why.

6.5.5 Picture task: was the robot perceived as social?

Children were asked to choose a first picture that best fitted with the robot/tablet, and explain their choice. They were then asked to choose a different second picture and explain their choice. Table 6.6 gives an overview of the pictures children chose in each condition.

In both conditions the majority of children first selected the laptop as the best fit for the robot/tablet. However, there was some more variation in the selection of pictures in the robot condition compared to the no-robot condition. Interestingly, some children in the robot condition chose the picture of the children, while no one in the baseline no-robot condition did.

Additionally, children were asked to select a first, and then a second picture that did not fit with the tablet/robot. Children in the robot condition most often chose the picture of the car or the dog as a picture that did not fit with the robot, as shown in Table 6.7. In the baseline no-robot condition most children chose the dog or the teddy bear as a picture that did not fit.

Children were asked to explain in their own words why they chose a specific picture. Their explanations were annotated on several labels that emerged from the
Now we’re talking: Learning by explaining your reasoning to a social robot

Table 6.6: Pictures that were chosen that did fit with the tablet or robot in the baseline no-robot and robot condition, respectively.

<table>
<thead>
<tr>
<th>Picture</th>
<th>No-robot 1st choice</th>
<th>No-robot 2nd choice</th>
<th>Robot 1st choice</th>
<th>Robot 2nd choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>0%</td>
<td>30.4%</td>
<td>12.5%</td>
<td>8.7%</td>
</tr>
<tr>
<td>Teddy bear</td>
<td>0.0%</td>
<td>4.3%</td>
<td>0.0%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Note book</td>
<td>8.0%</td>
<td>26.1%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Dog</td>
<td>4.0%</td>
<td>4.3%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Laptop</td>
<td>80%</td>
<td>0.0%</td>
<td>50.0%</td>
<td>26.1%</td>
</tr>
<tr>
<td>Teacher</td>
<td>8.0%</td>
<td>34.8%</td>
<td>16.7%</td>
<td>39.1%</td>
</tr>
<tr>
<td>Friends</td>
<td>0.0%</td>
<td>0.0%</td>
<td>20.8%</td>
<td>21.7%</td>
</tr>
</tbody>
</table>

Table 6.7: Pictures that were chosen that did not fit with the tablet or robot in the baseline no-robot and robot condition, respectively.

<table>
<thead>
<tr>
<th>Picture</th>
<th>No-robot 1st choice</th>
<th>No-robot 2nd choice</th>
<th>Robot 1st choice</th>
<th>Robot 2nd choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>9.5%</td>
<td>9.5%</td>
<td>30.4%</td>
<td>17.4%</td>
</tr>
<tr>
<td>Teddy bear</td>
<td>33.3%</td>
<td>52.4%</td>
<td>13.0%</td>
<td>30.4%</td>
</tr>
<tr>
<td>Note book</td>
<td>0.0%</td>
<td>4.8%</td>
<td>13.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Dog</td>
<td>47.6%</td>
<td>33.3%</td>
<td>30.4%</td>
<td>39.0%</td>
</tr>
<tr>
<td>Laptop</td>
<td>4.8%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Teacher</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Friends</td>
<td>4.8%</td>
<td>0.0%</td>
<td>13.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

data. Table 6.8 provides an overview of how often the answers were coded according to the annotation labels. These annotation labels were based on the topics children described in their explanations. Children in both conditions mainly focused on the technology or design of the robot/tablet when choosing a picture that did fit. For example, “The robot is controlled by a computer”. Interestingly, children who worked with the robot also talked about the social characteristics of the robot. For example, “he [the robot] is friendly”. Some of them talked about them teaching something to the system, indicating that they perceive the robot as being capable of learning. For example, “I am actually the teacher now”. Similarly, when explaining why a picture did not fit, children in the robot condition mainly focused on describing the social characteristics of the robot, for example “[the car] is similar to a robot, but it does not bring you company”. In the baseline no-robot condition, children mainly focused on describing the tool/function of the tablet and its social characteristics.

6.6 Discussion

In educational literature, several studies show that making your reasoning explicit by explaining something either to yourself or another person can contribute to deeper understanding and better learning [39, 42, 78, 168]. Therefore, prompting children to engage in explanatory behaviour is associated with better learning. The results of this study show that children verbalized for a longer duration to the robot. From this
Table 6.8: Frequency of annotation labels coded in participants’ explanations for choosing a picture that did or did not fit with the robot or tablet.

<table>
<thead>
<tr>
<th>Annotation label</th>
<th>Picture did fit</th>
<th></th>
<th>Picture did not fit</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No-robot</td>
<td>Robot</td>
<td>No-robot</td>
<td>Robot</td>
</tr>
<tr>
<td>Technology/design</td>
<td>48.0%</td>
<td>45.0%</td>
<td>30.4%</td>
<td>15.4%</td>
</tr>
<tr>
<td>Tool/function</td>
<td>28.0%</td>
<td>2.5%</td>
<td>34.8%</td>
<td>11.5%</td>
</tr>
<tr>
<td>Social characteristics</td>
<td>0.0%</td>
<td>20.0%</td>
<td>34.8%</td>
<td>53.8%</td>
</tr>
<tr>
<td>I learned from the system</td>
<td>24.0%</td>
<td>12.5%</td>
<td>0.0%</td>
<td>11.5%</td>
</tr>
<tr>
<td>I taught the system</td>
<td>0.0%</td>
<td>20.0%</td>
<td>0.0%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

we conclude that questions asked by a social robot yielded longer verbalizations than those asked by a baseline CAL system.

To gain a better insight in the content of children’s verbalizations, we looked specifically at the breadth and depth of their explanations. Firstly, regarding the explanation breadth, we found that: (a) when children verbalized longer in the observe phase, they mentioned more relevant facets in their explanations when explaining to the robot compared to the baseline no-robot CAL system; and (b) children touched upon more relevant facets in the observe phase when explaining to the robot compared to the baseline no-robot CAL system. More specifically, children in the robot condition mentioned the facets weight of pots and distance of pots significantly more often. These facets are crucial for explaining the tilt of the balance and should therefore be included in a good explanation of the working of the balance. In several instances, we saw that children who worked with the robot started explaining the working of the balance, without being prompted to do so, in the observe phase while we did not see this for children who worked with the baseline CAL system. This could indicate that children are more eager to explain to the robot how the balance works compared to explaining to the baseline CAL system. Secondly, regarding the explanation depth, we found that children drew more logical associations between the various facets of their explanations in the observe phase, showing a deeper level of deductive reasoning. From this we conclude that the CAL system extended with the robot was better able in motivating children to give more elaborate and detailed explanations, which, according to literature, is an indication of deeper learning.

There are several possibilities that might help explain why the robot yielded these results. By adding the robot, several aspects were introduced that were not present in the baseline CAL condition. First, the robot is physically present. Second, the robot displayed facial expressions. Third, the robot displayed interactive gaze. Fourth, the robot displayed life-like behaviours such as blinking its eyes and moving its mouth appearing as if he was speaking. Last, the robot was introduced with a background story. Each of these factors might motivate children to provide more elaborate explanations.

For example, [132] found that the physical presence and embodiment of robots help improve cognitive learning gains during tutoring and [170] found a positive effect of a physically embodied social robot on thinking aloud behaviour in children. Furthermore, [231] found that social gaze movements increase children’s perception of animacy and likeability. In addition, a robot that displays both social gaze and deictic gaze (which the robot in this study did) is perceived as more helpful and this might engage learners more. Furthermore, [169] found that the way a robot is introduced
Now we’re talking: Learning by explaining your reasoning to a social robot

(e.g. background story) impacts the way participants verbalize and explain concepts. Finally, facial expressions are known to impact how people perceive a robot and these facial expressions can complement or enhance an agent’s message [24].

Although we cannot draw conclusions about which of these factors contribute to the observed effect in a particular way, we did conduct exit interviews in order to gain insight in whether the children perceived the robot as more social. We asked them whether they could tell something about the robot/tablet, to uncover what they found most notable. In their answers children from both conditions talked most about the interaction with the robot/tablet. However, slightly more children in the robot condition talked about the embodiment of the robot (8 of 24) than children who worked with the tablet (5 of 25). In the robot condition they hereby focused on two things: that the robot moved, and the size of the robot. Interestingly, children did not mention the facial expressions the robot displayed. In addition, we asked children what they would tell other people (in this case their parents) about the experiment. In the robot condition, children again mentioned the embodiment of the robot (7 of 24) more often than children in the baseline condition (1 of 25). Children mentioned the fact that the robot could move, but they also mentioned the voice of the robot which they perceived as pleasant and similar to a child’s voice. In the tablet condition, none of the children mentioned the voice although this was the same voice as the robot. In addition to the exit interview, we conducted a picture task. Interestingly, 20.8% of the children who worked with the robot choose the picture with the friends as a first picture, and 21.7% as a second picture that did fit the robot. None of the children in the baseline condition chose this picture to fit the tablet. This seems to indicate that some children associate the robot with friends, while non of the children associate the tablet with friends. When explaining why they chose a specific picture (either that fit or that did not fit) children in the robot condition mentioned the social characteristics more often than children in the no-robot baseline condition.

The exit interview and picture task seem to indicate that children perceive both systems as somewhat social. Children in the robot condition focused more often on the embodiment of the robot where they noticed that the robot moved, the size of the robot and its voice. Also, children seem to associate the robot more often with ‘friends’ than the tablet and focused on the social characteristics more often when choosing a picture. Although it might be interesting from a theoretical perspective to disentangle the contribution of each of these factors, this is difficult in a real-life study, since an anthropomorphic robot that does not move or display (friendly) facial expressions might be perceived as unnatural or even scary. It seems however, that a robot that has such a collection of aspects motivates children to provide more elaborate explanations, which is an indication of a better understanding of the learning content.

6.7 Limitations

Although we conducted this study as thoroughly as possible, there are some limitations to the results presented here.

Firstly, the analysis of the logical associations between facets mentioned in the children’s explanations (i.e. the depth of their explanations) focused only on objectively observable explicit verbal associations. This limited us to draw conclusions regarding how children explicitly externalise their thoughts and reasoning processes. However, it may be that children also make these associations implicitly in their internal thought
processes, without expressing this verbally to the CAL system.

Secondly, regarding the interview, this study was done with Dutch speaking children. In the Dutch language it is common to refer to inanimate objects (such as the tablet) as “he” instead of “it”. It is therefore difficult to interpret such responses from the children as a measure for their social reference towards the tablet or robot. This also applies to the introduction of the robot at the start of the session. Although we used the male pronoun, the robot was not strongly introduced as being male. From previous experience we know that most children see this particular robot as male, even when it is introduced in a fully gender neutral way [27].

Thirdly, the children were interviewed by the experimenter directly after the session. It is possible that children saw the experimenter as someone with a higher social status and were therefore inclined to give socially desirable answers. Such a bias could explain the high number of positive answers to the first warm-up question: Did you enjoy it? To somewhat reduce the social desirability bias in follow-up questions we asked children to describe their experience in general, and how they would describe their experience to others. However, it remains difficult to say how well we succeeded in this. In general, the social desirability bias can sometimes be minimised by carefully selecting and training interviewers [155] or avoiding face-to-face interviews altogether [72, 155]. If the situation allows for it, the use of indirect questions [65] or neutral questions [155] can be advisable. We expect the picture task to potentially have a lower bias in that respect, since it had no apparent “right” or “wrong” answers. Furthermore, it can be advisable to include a social desirability scale to measure and correct for an individual’s tendency to be affected by this bias [121]. Several scales have been developed specifically for children (e.g. [46, 147]).

Finally, to make more definitive statements about the perceived social nature of the system, it is important to improve the sensitivity and validity of the measurement tools. Developing such tools specifically for young children demands an iterative approach. The study reported in this chapter described the first iteration in the development of such a tool, consisting of a semi-structured interview and accompanying picture selection task. Since then, based on the insights from this first iteration, a second iteration of the tool has been designed and validated to measure perceived social competence of a robot, as described in [35].

6.8 Conclusion

This chapter presented a study in which children carried out an inquiry learning task with either a baseline CAL system or a CAL system extended with a social robot. Children were asked to verbalize their hypotheses, observations, and conclusions out loud to the system. The main research question was: *How do feedback and questions delivered through the RECAL system affect children’s explanatory behaviours?* Our expectation was that a robot would be more effective in eliciting verbalizations, because it is more intuitive to explain something to a social partner and the robot is a “social” entity. Results show that:

- Children verbalized longer when they explain to a robot compared to a baseline CAL system.
- Children who verbalized longer touched upon more relevant facets (explanation breadth)
• Children who worked with the robot made more logical associations between the relevant facets in their explanations (explanation depth)

These results show that extending a CAL system with a social robot can have a positive effect on children’s explanatory behaviour and might therefore have potential for influencing children’s learning behaviour in new and positive ways. We argue that in this situation the robot is able to make a meaningful impact on the child's learning process due to its nature of being a robot. Because it is a robot, it can leverage a rich social behaviour repertoire through which it may naturally express certain types of interactions, such as reactive and directive gaze and facial expressions. Additionally, due to it being a social agent, we can naturally position the robot in a certain role and present it to the user with a certain background story to match its behaviours.

The effect that a robot has on the social act of eliciting explanations can be studied over a relatively short period of time, in this case during a single session. However, beneficial effects of other social acts may only manifest after prolonged periods of repeated interactions. In the next chapter we present a longitudinal study with repeated sessions that explored how the social act of a robot giving praise affected children’s attitudes towards learning.
“Wow, that actually went very well!” Chucky feels like he now has a better idea of the challenge he's up against. “I've learnt a lot from all their explanations, but I wish there was more I could do.” He knows he'll need to up his game if he is to stay in the class permanently.

Mees walks up to the table and sits down with a sigh. “I'm bored!” he mutters, “I want to do a different task.”

“Well hello Bored,” he says with a smirk, “I'm Chucky!” “We're in luck, there's another task over there with marbles.”

“Ugh, that looks hard! I bet I suck at it!”

“Sounds like someone could use a healthy dose of motivation!”

7.1 Introduction

In the previous chapter we explored a setup where children offered verbal explanations to the system while working on a learning task. Using the RECAL system we investigated the effects of a robot carrying out the social act of eliciting explanations. Results showed that the robot had an impact: children gave better explanations when prompted by the robot than when prompted by the regular CAL system. Evidently, interactive multimodal technologies like robots can be very useful tools in education. The effects of social acts like prompting for explanations can be measured during

The work presented in this chapter was carried out together with Frances Wijnen and has been submitted as D. P. Davison et al. Words of encouragement: How praise delivered by a social robot changes children’s mindset for learning. Journal on Multimodal User Interfaces, 15(1):61–76, 2020.
relatively brief, self-contained interactions (e.g. a single session). However, first-time
users who are not familiar with robots may show different behaviour during such brief
interactions than they usually would. As users become accustomed to the robot and
the interaction setting these novelty effects begin to wear off. Once this happens, we
can start to see how users might interact with robots after prolonged use.

This is an important distinction because some types of social acts may only show
effects after repeated interactions over extended periods of time. To investigate
whether a robot can also have an impact here, we must move towards longitudinal
studies. Inspired by the findings from the previous short-term study on verbalisation
we investigated other social aspects of learning where robots might be able to enhance
the delivery of the social acts of the CAL system. The social act of giving praise being
one such instance where we saw opportunities for a social robot to play a meaningful
role. This chapter presents the results of a long-term, unsupervised, in the wild study
where we investigated effects of praise by a CAL system, delivered through a social
robot, on children’s attitudes towards learning.

During this study we also collected data about how children worked with the system
and progressed through tasks, and how this evolved over time. Those exploratory
results are presented in Chapter 8.

7.2 Background

Praise has long been recognised as an important social mechanism that can be used to
support a learner. Praising the learner’s process, abilities, and achievements has been
shown to influence their motivation, performance, and self-esteem, among others [74,
152, 95, 109].

One of the factors that plays an important role in the motivation for learning and
thinking in school settings is a learner’s mindset [25]. Dweck [60] describes two
forms of mindset: 1) a fixed mindset is characterised by the belief that you are born
with a certain capacity and that you cannot influence your capacity very much; 2) a
growth mindset is characterised by the belief that you can improve your capabilities
and expertise through perseverance and effort, and that failure is an inherent part of
learning.

On the one hand, Mueller and Dweck [152] and Dweck [60] show that people
with a fixed mindset tend to focus on proving their intelligence: they want to look
smart. Because of this, they are reluctant of tasks that are challenging or hard, because
there is a chance of failure, which conflicts with their goal of looking smart. In their
belief, failing at something is an indication that you are not smart enough or lack the
necessary capabilities; when you fail it is an indication that you will not be able to
complete a certain challenge. You better give up and try something easier. Further-
more, according to people with a fixed mindset, effort is seen as something negative.
If you have talent or are gifted then it isn’t necessary to struggle with a task. Viewed
from such a perspective effort is for people who lack talent.

On the other hand, people with a growth mindset tend to focus on learning. They
are motivated to do new and complicated tasks because it provides opportunities for
learning. Their goal is to learn and to develop themselves by working hard and by
putting in a lot of effort. They see failure as something that is necessary for learning.
So after failing, people with a growth mindset tend to work harder and try out new
strategies in order to complete a challenge or master a skill [61, 60, 152]. A growth
mindset is seen as a favourable trait when it comes to exploring new learning domains and developing new skills. Therefore, we focused on promoting such a growth mindset.

Praise and criticism have an important influence on the development of growth and fixed mindsets [74, 152, 95]. For praise to have a positive impact it is important that it is perceived as contingent, specific, sincere and credible [153, 158]. Praise for high ability and personal traits is a common response when someone did a job well. Whether it is in the classroom, when playing sport, or during artistic endeavours, praise for ability is seen as a popular tool to stimulate learners’ motivation [152, 109]. However, focusing on primarily praising high ability may have an undesired impact. It can make children feel pressured to perform well in future situations, which stimulates a fixed mindset. An alternative for praising ability is praising effort. Instead of praising one’s goal to seem smart, effort-related praise focuses on the process of learning or mastering of a certain skill [156]. This form of praise stimulates a growth mindset, since the emphasis is on the process of learning instead of the end result [152].

In related work involving robots and mindset, Park et al. [164] have shown that a peer-like robot can promote a growth mindset in children. The robot and child took turns solving puzzle tasks, during which the robot either exhibited neutral behaviours or role model behaviours associated with a growth mindset. Their robot used a multimodal behaviour repertoire consisting of speech, nonverbal expressions, and gaze. Depending on the condition, the robot would use neutral factual statements or mindset-related statements accompanied by an appropriate body posture and facial expression (e.g. engagement, interest, excitement, or frustration). Afterwards, children who had worked with the role model robot exhibited a stronger growth mindset themselves. In our study, in contrast with Park et al. [164], we were interested in investigating the role of praise instead of role modelling behaviour.

7.3 Aims and objectives

This chapter aims to gain deeper insights into how children respond to effort-related praise while working on a learning task. Offering praise to a learner can be seen as a social act of the CAL system. We are interested in exploring ways in which a social robot may be used to extend a traditional CAL system to deliver such social acts in a natural and convincing way. In situations where a child works together with a peer learner robot, we consider praise to be an appropriate and positive way of supporting learning, more so than criticism (even if constructive). Therefore, we chose to primarily focus on promoting attitudes associated with a growth mindset as opposed to dissuading attitudes related to a fixed mindset.

This leads to the following research question: What are the effects on the mindset of children when extending an autonomous CAL system with a social robot to deliver effort-related praise?

We expect that the social act of giving effort-related praise has a greater potential impact when it is delivered by a robot as opposed to a traditional CAL system, since by its very nature a robot can be more convincingly presented as a social entity. A robot has an elaborate repertoire of social cues to engage with the learner, such as focus of attention, facial expressions, and deictic gaze. Therefore the hypothesis in this study was: Participants who work with a CAL system that delivers praise through a social robot will display a stronger growth in their mindset than participants who work with a CAL system that delivers such praise without a social robot.
Additionally, children from a different school worked with a baseline version of the interactive CAL system, without robot, which offered no such effort-related praise. Although for these children we found no significant effects on their mindset, for the sake of completeness we also report these results of the baseline CAL system here.

7.4 Design of system and multimodal interactions

This study used Variant IV of the RECAL system as described in Section 5.6. It was based on a technical architecture that was developed as part of the EASEL project [171], which was adapted and further extended to support fully autonomous, unsupervised, long-term interactions in the wild. An early prototype of this system using interactive embodied learning materials has been previously described in [48]. The system consisted of the following main components: embodied learning materials, sensors, headphones, a tablet, a robot, and a control computer.

The embodied learning materials were based on learning instruments originally developed by Inhelder and Piaget [82] to support the exploration of several phenomena from the physics domain. Building on principles of inquiry learning these tools are designed to support a scientific process of discovery [105, 104]. In the first task, adapted from our earlier work [53, 220], children used a balance scale to explore the moment of force. They placed combinations of three differently weighted pots on three distances from a central fulcrum to discover how those variables affected the tilt of the balance. The balance task is shown in Figure 7.1. The second task, newly developed for this study, consisted of two sloped ramps from which children could race balls to explore potential energy and rolling resistance, while the angle of each slope could be adjusted to a low or high position. The children could select balls from a combination of materials, weights, and densities: small marbles; medium pingpong balls, wooden balls, and rubber balls; and large styrofoam balls and marbles. By racing them head-to-head, they discovered how their speed was influenced by these variables. The ramp task is shown in Figure 7.2.

Several different types of sensors were used to enable fully autonomous unsupervised interactions. Firstly, embedded sensors in the learning tasks recorded the state of the materials and the child’s actions. The balance scale task used a potentiometer in the central pivot to measure the tilt of the balance, it used different resistors in each pot to measure their placement on the various locations, and it used reed switches to detect whether the stabilising blocks were present under each side of the balance. The ramp task used potentiometers to measure the angle of each slope, it had a physical button for releasing the balls, and it used a photoresistor to measure when a ball had reached the end of the track. Secondly, an external Microsoft Kinect\footnote{Microsoft Kinect for Windows v2: https://support.xbox.com/en-US/xbox-on-windows/accessories/kinect-for-windows-v2-info} depth-camera was used with the SceneAnalyser software [232] to detect the presence of a child and the location of their face. Finally, an RFID scanner was used to recognise individual children; we handed out unique RFID badges that children scanned at the start of each session.

The system supported several forms of multimodal inputs and outputs. A Samsung Galaxy Tab2 tablet was used as the primary means of direct input to the CAL system: children could input their responses to questions by pressing buttons on the tablet interface. The various sensors in the task provided another means of input to the
Figure 7.1: Setup of the balance scale task with the following components: A) balance scale learning materials with embedded sensors; B) tablet; C) headphones; D) Microsoft Kinect and wide-angle webcam; E) RFID card scanner; and F) Zeno R25 robot.

system, enabling it to respond dynamically while children worked with the task.

Task instructions were delivered verbally and displayed on the tablet as written text and illustrations (shown in Figure 7.3). Since the study took place in a real classroom during school time, teachers had requested that interacting with the system should not interrupt or disturb regular lessons. Such requests are common when doing a study in class [112]. Therefore, all verbal utterances produced by the system and robot were played through headphones. The system used the Fluency1 text-to-speech engine to generate Dutch speech.

Robokind’s Zeno R252 robot was used in the Robot condition to deliver the CAL system’s verbal feedback and praise. This small humanoid robot has a face that can express basic emotions. Furthermore, it has several degrees of freedom in its eyes,

1Fluency text-to-speech engine for Dutch speech: https://www.fluency.nl/tts/
2Robokind: https://www.robokind.com/
The robot’s multimodal behaviours in this study were informed by design guidelines emerging from an extensive contextual analysis of inquiry learning tasks with our target user group [53], previously described in Chapter 4. The robot used expressions for smiling when addressing the child and delivering praise, and amazement when the child performed the experimentation step of the task (for example, see Figure 7.4). Furthermore, it gazed dynamically towards the user, the tablet, and relevant areas of the task depending on the child’s actions and progress (for example, see Figure 7.5). The robot used lip synchronisation to match the verbal utterances produced by the text-to-speech engine. Minimal idle behaviour was added through periodic eye blinking.

All control software ran on a small desktop computer, which was hidden out of
Figure 7.3: Assignment instructions displayed on the tablet during the preparation phase for the balance and ramp tasks. Each assignment had both a textual and a visual description. Children could press the bottom left button to have the assignment text read out loud. By pressing the bottom right button, children continued to the next assignment phase: prediction.

sight of the children. This computer was responsible for collecting and interpreting the sensor values and generating responses from the system. An early version of Flipper 2.0 [215] was used to define the CAL system’s dialogue models, and to manage the flow of the interaction. Behaviours of the CAL system were specified in Behaviour Markup Language (BML) [110], which were executed by ASAPRealizer [172]. ASAPRealizer is an engine for choreographing multimodal behaviours across devices, modalities, and platforms. It ensures that the verbal utterances, tablet interface updates, and robot movements remain synchronised throughout the interaction. The core BML specification was extended with robot-specific and tablet-specific behaviours that could not be expressed in standard BML: gazing towards a $(x, y)$ location and showing text, images, and buttons on the screen.

An example of a multimodal BML behaviour script is shown in Figure 7.6. The timing of speech, gaze, facial expressions, and content displayed on the tablet are synchronised to produce a coherent behaviour sequence. Placeholder variables, like the assignment details and the location of the child’s face, are filled at runtime based on the available sensor data and the child’s progression through the tasks. In this example the tablet displays an assignment including text and an image. When the content appears on the tablet, the robot smiles, shifts his gaze from the child towards the tablet and starts reading the assignment text out loud. When finished reading, the robot shifts his gaze back to the child.
Figure 7.4: Zeno's facial expressions.

(a) Smiling

(b) Amazed

Figure 7.5: Zeno dynamically shifts his gaze at various moments during the interaction.

(a) Gazing towards the user. The location of the user was detected by the SceneAnalyzer software [222].

(b) Gazing towards a part of the task. The gaze direction and the appropriate moment for gazing was determined using embedded sensors in the task (here, the balance had just tilted to one side).

7.5 Methodology

7.5.1 Study design

The main study was a between-group design with one independent variable: the presence or absence of a social robot to deliver the system’s effort-related praise. The dependent variable was the child’s mindset, measured through a pretest and posttest questionnaire.

7.5.2 Conditions

We manipulated whether the CAL system’s praise was delivered by a robot or not. This resulted in the following two conditions:
Words of encouragement: How praise delivered by a social robot changes children’s mindset for learning

Custom BML extensions for supporting the robot and tablet

Fine-grained synchronisation between behaviours

Various placeholders are filled at runtime

Figure 7.6: A BML script illustrating multimodal behaviours of the system when reading an assignment out loud.

- **[No-Robot]** a traditional CAL system that offered effort-related praise
- **[Robot]** a CAL system extended with a robot to deliver the effort-related praise

Over the course of several months children worked individually on two consecutive learning tasks (see Figures 7.1 and 7.2) that consisted of assignments of increasing difficulty. In both conditions, the CAL system offered identical task-related instructions, help, feedback, and effort-related praise.

### 7.5.3 Participants

Both conditions were run in parallel in a classroom of two district locations of the same Montessori school, located in a comparable suburb of the same city. The demographics
Table 7.1: Participant demographics. In the Montessori and Freinet educational systems children of mixed age groups learn together in the same classroom.

<table>
<thead>
<tr>
<th></th>
<th>Age Mean (SD)</th>
<th>Nr. of participants</th>
<th>Total (girls/boys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Montessori No-Robot condition</td>
<td>7.1 (0.83)</td>
<td>24</td>
<td>10/14</td>
</tr>
<tr>
<td>Montessori Robot condition</td>
<td>7 (0.82)</td>
<td>20</td>
<td>8/12</td>
</tr>
<tr>
<td>Freinet Baseline CAL system</td>
<td>8.8 (0.73)</td>
<td>17</td>
<td>12/5</td>
</tr>
</tbody>
</table>

of these classes were very similar, allowing for cross-sample comparisons; children in both classes were of the same age range and there was a similar ratio of girls and boys. Participants were 44 children between 6-10 years old, as described in Table 7.1.

In parallel, the baseline CAL system without robot and without praise was tested with 17 children in a classroom of a Freinet school of the same city. The demographics of the Freinet class were different than the Montessori classes; on average these children were almost two years older and the gender ratio was reversed. This makes it more difficult to draw cross-sample comparisons between the Freinet and Montessori school classes for the scope of this study.

Ethical approval was obtained from the EEMCS ethical board of the University of Twente and parents signed an informed consent letter prior to the start of the study. Prior to this study it was agreed with the school directors that the learning tasks would fall under the school’s regular science education curriculum. Children who did not obtain consent from their parents still had the option to work with the learning tasks in order to benefit from the educational content. However, in these cases the system collected no data and we conducted no interviews with these children.

7.5.4 Manipulations: delivery of praise

Both conditions used the same state-of-the-art CAL system which offered verbal task-related instructions, task-related help, task-related feedback, and effort-related praise using a computer-generated voice. To not disturb and distract other children in the classroom, participants always used headphones to listen to the system’s verbal utterances.

In the No-Robot condition, children worked with the CAL system without a robot, and the praise was delivered only through the headphones. In the Robot condition, the CAL system was extended with a social robot to deliver the same effort-related praise. The robot would gaze towards the user and show a smiling facial expression while verbally delivering the praise. Both the No-Robot and Robot conditions offered identical praise using the same computer-generated voice, played through the same headphones.

The system offered such praise at several moments during the assignments. Firstly, after a child completed the experimentation phase of an assignment and had entered their observation the system would offer a compliment on their progress, such as “I think you put in a lot of effort!” or “I see you tried your best!”.

Secondly, after the conclusion phase, the system would ask them how they felt about whether their hypothesis was correct or incorrect. For example, if their hypothesis was incorrect, they could select either “I think I am not good at this” or “I think I can learn this” on the tablet. After selecting the former, the system would respond with “We can
learn from a mistake!”. After selecting the latter, the system would respond with “I think so too!”.

Finally, after each completed assignment the system asked the child to rate how difficult they found the task, after which they could choose the difficulty of the next task. At this point, the system would generate appropriate feedback and praise to promote a growth mindset. The praise that was given by the system depended on three aspects: 1) whether or not the child gave a correct hypothesis; 2) the self-reported assignment difficulty; and 3) the subsequent selected level of difficulty. Based on these aspects the system labelled the child’s attitude at that point in time as either performance-driven or mastery-driven.

On the one hand, a performance-driven attitude is characterised by wanting to demonstrate competence by avoiding mistakes, thus often shunning difficult or unknown challenges. Individuals with a fixed mindset often exhibit performance-driven attitudes. When a child exhibits such performance-driven behaviours the system offered feedback to help promote a growth mindset. For example, when a child made a correct prediction, indicated that they found the assignment easy, yet still chose an easier next assignment, this was labelled as performance-driven. In this case, the system would highlight the importance of seeking an adequate challenge by giving the feedback “Gosh, I would have expected you to choose a more difficult task, because then you can learn more” or “It is fine if you want to practice more, but if you want to learn something new we can try a more difficult task”.

On the other hand, learners with a mastery-driven attitude focus on improving their skills and learning process through practice, thus often embracing more difficult challenges. Such attitudes are often associated with individuals who lean towards a growth mindset. When a child exhibits such mastery-driven behaviours the system offered feedback to further strengthen their growth mindset. For example, when a child made an incorrect prediction, indicated that the difficulty was okay, and chose the same difficulty for their next assignment, the system labelled this as a mastery-driven attitude. In this case, the system offered encouragement to emphasise the importance of practising: “We didn’t get it yet this time, let’s practice some more and we can learn!”. Similarly, for example, if a child made a correct prediction, indicated that the assignment was easy, and chose a harder next exercise, the system would praise the child to emphasise the importance of seeking a challenge: “Great, you are choosing a challenge, I like that!” or “Great, during a more difficult task we might learn something new!”

7.5.5 Measures

The main research question in this study focused on affecting the mindset of children. In particular, we were interested in promoting a growth mindset through effort-related praise. Similar to Park et al. [164], pretest and posttest questionnaires were used to measure a change in the children’s mindset as a result of the intervention. The 18-item questionnaire used in this study was inspired by the questionnaire designed by De Castella and Byrne [54] who revised the implicit theories of intelligence scale designed by Dweck [59]. Colleagues from the ELAN group of the faculty of Behavioural, Management and Social Sciences of the University of Twente used a part of the questionnaire from De Castella and Byrne [54] to design a version that is suitable for young children.

In contrast with the questionnaire presented by De Castella and Byrne [54], the
concept “intelligence” was replaced by “smart”, as pilot tests showed that this was better understood by very young children. The following (translated) definition of smart was given to the children: “Smart means that you are well able to consider, think up, and thrash out/figure out.” The questionnaire consisted of items that fall under two main constructs: items measuring a growth mindset and items measuring a fixed mindset. Furthermore, additional questions regarding effort were added to the questionnaire used in this study, since beliefs about effort are related to mindset.

Children could provide their answers according to a 4-point Likert scale with the following options: strongly agree, somewhat agree, somewhat disagree, and strongly disagree. However, Likert scales are often difficult for very young children because they tend to think more dichotomously and have a tendency to endorse responses at the extreme end of the presented scales. This can especially be the case if the statements are related to more ‘fuzzy’ subjects such as feelings, beliefs or attitudes [145]. Park et al. [164] addressed this by offering children sets of bipolar statements from which to choose. However, we were interested in capturing nuanced responses of children that were not necessarily on either extreme end of the spectrum.

In several pilot test iterations with our target user group, we explored different techniques for administering the questionnaire. Participants in these pilot tests were from several primary schools visiting the university during school trips, for whom signed parental consent was available. Firstly, we presented the questionnaire in the traditional fashion as a self-administered test, while giving children the option to ask the experimenter for help or additional explanation. Only some of the older children were able to complete this version of the questionnaire without any issues, as the younger children had difficulties reading and understanding the questions. Secondly, we had the experimenter read each statement of the questionnaire out loud, after which children were asked whether they strongly agreed, somewhat agreed, somewhat disagreed, or strongly disagreed with the statement. In this case, some children seemed to be unable to distinguish between the answer options or were hesitant to commit to a choice of answer. Finally, the technique which was best understood by the target group was to first have the experimenter read each statement aloud and then ask: “do you agree or disagree?” We observed that children could quite naturally answer this dichotomous question. After a child made an initial choice, the experimenter would subsequently ask: “do you strongly (dis)agree or somewhat (dis)agree?”

The resulting mindset questionnaire was implemented as an interactive web form and used as a pretest and posttest. The following items were used (translated from Dutch):

1. I think I can change how smart I am
2. I think I can’t change how smart I am, because I am born like this
3. I think I will always stay this smart, because I can’t change that
4. I think I can become smarter step-by-step
5. I think I will always stay this smart, because that is fixed in my brain
6. I think I can change how smart I am by practising with assignments of increasing difficulty
7. I think it is fixed how smart I am and there is nothing I can do to change that
8. I think I can change how smart I am, by doing my best

9. I think that my smartness is fixed in my brain, and I can’t change it

10. I think I can always change how smart I am

11. I work harder on difficult assignments because then I learn the most

12. I feel dumb when I have to think really hard for an assignment

13. I do my best for difficult assignments, because then I learn the most

14. I do my best for difficult assignments, because then I will be able to do them better

15. I work much less hard for difficult assignments, because I can’t do them anyway

16. I work less hard for difficult assignments because I prefer not to do much effort

17. I prefer to choose more difficult assignments, because then I can learn something new

18. I prefer to choose easier assignments, because then I have to spend less effort

7.5.6 Procedures

Due to the long-term nature of this study, we worked together with the teachers and school management to fit our activities in with their regular school schedule as best as possible. In some cases, this meant that activities at individual schools were moved forward or delayed and that schools followed a slightly different timeline throughout the study. Additionally, the total duration of the experiment varied per school due to holidays and other events. Table 7.2 shows an overview of the stages of this study for each school, highlighting the timing of questionnaires and tasks.

The study took place between December 2016 and May 2017. Around two weeks before the start of the study two experimenters would do the mindset pretest with the children. The mindset posttest was done in the week after the second task ended. The experimenters would call each child one by one to a separate room in the school. The experimenter then read aloud the statements of the mindset questionnaire and noted down the answers of the children. It took children approximately 10 minutes to complete the pretest and posttest. Following the pretest, the first learning task was placed in the classrooms for approximately 6-7 weeks. During this period children initiated a total of 260 sessions and completed a total of 756 assignments. Then, the second task was placed in the classrooms for approximately 8-10 weeks, during which a total of 195 sessions were initiated and children completed a total of 550 assignments. In all three classes children progressed through the various levels of difficulty without many issues, with the majority of children achieving the highest level in each of the two tasks at some point. More detailed procedures for each task are discussed in the following sections.
Table 7.2: Timeline for conducting the various stages of this study in each of the three participating schools. The duration mentioned for each task excludes weekends and holidays. In the No-Robot condition children worked with a CAL system that offered praise. In the Robot condition, this praise was delivered by the robot. Additionally, we tested a Baseline CAL system without a robot and without praise.

<table>
<thead>
<tr>
<th>School location 1</th>
<th>School location 2</th>
<th>School location 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No-Robot condition</strong></td>
<td><strong>Robot condition</strong></td>
<td><strong>Baseline CAL</strong></td>
</tr>
<tr>
<td>Pretest questionnaire</td>
<td>16/12/2016</td>
<td>15/12/2016</td>
</tr>
<tr>
<td>Task 1 - Balance</td>
<td>9/1/2017 - 24/2/2017</td>
<td>10/1/2017 - 24/2/2017</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td>30 schooldays</td>
<td>29 schooldays</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td>36 schooldays</td>
<td>35 schooldays</td>
</tr>
<tr>
<td>Posttest questionnaire</td>
<td>30/5/2017</td>
<td>29/5/2017</td>
</tr>
</tbody>
</table>

7.5.6.1 Task 1 - Balance

The balance task (see Figure 7.1) was placed in the classroom after school hours. The next school day an experimenter gave a short explanation to all children, introducing the various components of the system: the tablet, headphones, the learning materials, and the robot. Children were instructed that the voice of the system/robot was speaking through the headphones. They then handed out the personal RFID badges and showed children how to scan their badge to initiate the interaction. The children were not instructed how often or how long they should interact with the learning task, and there was no set schedule. Instead, children were entirely free to work with the system on their own initiative, as long as it fit within the lesson schedule of the teacher.

To initiate a learning session the child would put on the headphones and scan their RFID tag. The system would then greet the child by speaking through the headphones. If it was the very first interaction, the system would greet the child by saying: “Hi [NAME], nice to meet you! Shall we play together?” If the child had already interacted with the system before the system would say: “Hi [NAME], it’s nice to see you again. How did it go last time? Do you think it was hard, easy, or was it okay?” [CHILD SELCECTS ANSWER] “All right, last time we finished assignment [DIFFICULTY LEVEL] let’s move on from here.” After every task, the child would choose whether they wanted to do another task and whether the next task should be easier, harder, or equally difficult.

If the child indicated that they did not want to play any more, the CAL system said goodbye and logged out the current user: “Goodbye [NAME], it was nice playing with you! See you next time!” If the child indicated that they wanted to continue, the system started a new task according to the chosen difficulty level. A child could complete a maximum of four assignments during each interaction. After four assignments the system said goodbye to the child and logged them out automatically. It took children approximately 10 minutes to complete four assignments.

7.5.6.2 Task 2 - Ramp

Similar to the first task, the ramp (see Figure 7.2) was placed in the classroom after school hours. The next school day, a researcher gave an explanation of the new task. Since at this point children were familiar with the tablet, headphones, and robot,
the explanation focused on introducing the new task. In contrast with the highly predictable deterministic nature of the balance assignments, the ramp task would sometimes give unpredictable results: balls would occasionally bounce off the sides of the ramp while rolling, which would cause them to slow down. Especially when racing two otherwise identical balls, this occasionally resulted in an unexpected difference in their finishing times. Children were therefore reminded and encouraged that they could repeat the same race as many times as they wanted, in order to collect additional evidence to confirm or refute their initial observations.

Children again used their RFID badge to start sessions with the system on their own initiative. The procedure, the system utterances, and praise mechanism were the same as in the first task. Depending on the difficulty level and how often children would repeat the same experiment, they spent between 5-15 minutes on completing four assignments.

7.6 Analysis

In the scope of this chapter we were interested in the effects on the mindset of children when extending an autonomous CAL system with a social robot to deliver the system's effort-related praise. An analysis of how children worked with the system are reported in Chapter 8; here we focus on the analysis of the mindset questionnaires.

The mindset questionnaire was used to gather pretest and posttest scores for participating children. Answers were marked according to a 4-point Likert scale with the following options: strongly disagree, somewhat disagree, somewhat agree, and strongly agree. Answers were then converted to a numerical representation by assigning the respective scores 1, 2, 3 and 4, such that a low score corresponded with disagreement and a high score corresponded with agreement.

To investigate the presence of underlying constructs in the questionnaire items an Exploratory Factor Analysis (EFA) was performed using all pretest questionnaire scores from the three participating classes. Bartlett’s test of sphericity shows that the data is appropriate for EFA, $\chi^2(153) = 291, p < .001$. We used an oblimin factor rotation and a factor loadings cut-off value of 0.4. Remaining items with a loading below 0.4 on all factors have been dropped from further analysis. Following this approach, three mutually exclusive constructs emerged from the data:

**Growth** The ‘growth’ construct consists of items 1, 4, 8, 13 and 14 which explain 41% of the variance, with loadings ranging from 0.46-0.72. These items are formulated in such a way that they align with a mindset oriented towards growth. They address that one can change how smart they are, and that one can learn and do better by working hard on difficult assignments. When a participant shows high agreement with these items this corresponds with a growth mindset. A Cronbach’s Alpha of 0.73 shows an acceptable internal consistency for this construct.

**Fixed** The ‘fixed’ construct consists of items 2, 3, 5, 7 and 9 which explain 31% of the variance, with loadings ranging from 0.41-0.77. These items highlight one's inability to influence how smart they are. The items cite innate causes for this lack of influence. This is an attitude that is characteristic for individuals with a fixed mindset. Accordingly, when a participant shows a high agreement with
these items this corresponds with a fixed mindset. A Cronbach’s Alpha of 0.68 shows a questionable internal consistency for this construct.

**Effort** Finally, the ‘effort’ construct consists of items 16 and 18 which explain the remaining 28% of the variance, with respective loadings of 0.99 and 0.43. These two items describe a preference for spending less effort, either by working less hard on difficult assignments or by choosing easier assignments, both of which are telling of a fixed mindset. When a participant shows a high agreement with these items this corresponds with a fixed mindset. A Cronbach’s Alpha of 0.59 shows a poor internal consistency for this construct. Therefore, this has been dropped from further analysis. Future iterations of this questionnaire should expand the number of relevant items, and may focus on further validation of the construct as a whole.

Only participants who completed both the pretest and posttest were included in the data set. A score was computed for each construct by taking the average of the individual items belonging to that construct, such that a low average score corresponds to a low agreement with the statements in that construct and a high score corresponds to a high agreement.

### 7.7 Results

A total of 17 children worked with the baseline CAL system, which we tested in a separate school in parallel with the main study. For these children we found no significant differences between their pretest ($M = 3.5, SD = 0.4$) and posttest ($M = 3.4, SD = 0.5$) scores.

Regarding the main study, a total of 24 and 20 children completed both the pretest and posttest in the No-Robot and Robot conditions, respectively. Pretest and posttest results for the growth and fixed constructs are shown in Figure 7.7. We found no significant differences between conditions for the fixed construct. For the growth construct, a repeated measures ANOVA showed no significant interaction effects, but did show significant main effects on the between-subjects variable (the source of the praise) (Repeated measures ANOVA, $F_{1,42} = 7.23, p = .01$) and the within-subjects variable (the pretest versus posttest scores) (Repeated measures ANOVA, $F_{1,42} = 4.7, p = .036$). These results show that there was a difference between the conditions and that during the course of the study the feedback and praise led to an overall increase in growth mindset. However, these results do not necessarily show that this increase was mediated by the robot.

To further investigate the growth construct main effects, a post hoc analysis was performed with Bonferroni corrected pairwise tests (adjusted $\alpha = .0125$). Between subjects, we found no significant difference on the pretest scores for the No-Robot ($M = 3.2, SD = 0.7$) and Robot ($M = 3.4, SD = 0.5$) conditions. However, we did find a significant difference on the posttest scores between the No-Robot ($M = 3.3, SD = 0.5$) and Robot ($M = 3.8, SD = 0.2$) conditions (Wilcoxon rank sum test, $U = 376.00, p = .0011$). Within subjects, we found no significant differences in the No-Robot condition between the pretest ($M = 3.2, SD = 0.7$) and posttest ($M = 3.3, SD = 0.5$) scores of the growth construct. However, in the Robot condition we did find a significant increase from the pretest ($M = 3.4, SD = 0.5$) to the posttest ($M = 3.8, SD = 0.2$) scores (Wilcoxon signed rank test, $z = -3.19, p = .0014$). These results
Words of encouragement: How praise delivered by a social robot changes children’s mindset for learning

Figure 7.7: Pretest and posttest results for the growth and fixed constructs of the mindset questionnaire. Repeated measures ANOVA showed significant main effects in the growth construct on the source of the praise and on the pre and posttest.

show that the group of children who received praise from the robot saw a significant benefit to their growth mindset.

7.8 Discussion

Although the results do not necessarily show that an increase in growth mindset scores was mediated by the presence of the robot, they do show that the CAL system as a whole has had a positive effect. Furthermore, we found a significant improvement in the growth mindset in the Robot condition, where no significant difference was found in the No-Robot condition. Based on these results we interpret that working with and receiving praise from the robot had a positive effect on children, although this study was unable to uncover exactly what caused this.

Working with a robot may have all kinds of impact on how children work with an interactive learning system. In previous work we looked at effects on explanation behaviour [220] and in addition to the mindset results reported here, we looked at how children’s interactions with the system developed as they progressed from initial novelty effects towards sustained use [51]. Results from the study presented in this chapter and related work from Park et al. [164] suggest that robots who offer implicit or explicit remarks, feedback, and praise may promote a growth mindset. Promoting such a growth mindset in young learners has been shown to improve academic achievement later in life [23], which leads us to speculate that having a robot in class can potentially have a positive impact on learning in the long run. This line of research suggests that robots can be promising tools for education.

Other research has shown benefits of promoting a growth mindset with older learners. First year high school students, for instance, have been found to significantly improve their grades after participating in a single-session mindset intervention [226]. Similar interventions may continue to help underachieving students throughout high school, increasing their performance and grades, potentially lowering the chances of them dropping out [165]. Besides learners, teachers may also benefit from mindset interventions. Seaton [186] found that teachers not only improved their own mindset, but could also more confidently apply it in practice in their own teaching. It could be
an interesting line of future research to investigate how a robot’s mindset support may adapt, grow, and evolve with the needs of learners of all ages.

In pursuit of ecological validity we conducted the longitudinal study unsupervised and in the wild. By doing so, we have shown that it is feasible to conduct comparative HRI studies in real classrooms, capturing real changes in children’s learning process. As a consequence, however, we identified two main limitations related to this study. Firstly, since this study spanned a relatively long period of time we were unable to follow identical procedures and timelines in all participating schools, despite our best efforts. Special events such as holidays, sports days, and school musicals took place at different moments in the experiment timeline for the different schools. While we do not expect this to have impacted their mindset directly, it may have had an influence on the number of sessions children initiated throughout the study. Secondly, although we took care to select similar schools from a similar region, the results of this study may have been impacted by differences in educational methods between schools, lesson plans from individual teachers, or other external factors. That being said, the schools’ curricula did not explicitly cover the topic of mindset, and in discussions with the teachers we found no indications that would suggest major differences between classes.

To prevent distractions in the classroom, teachers had requested that the robot would not make too much noise. Therefore we did not use the robot’s rich full-body behaviour animation repertoire, resulting in a somewhat static experience. When talking with children after the study, some were disappointed in the robot’s limited movements (e.g. “he didn’t move his arms or legs” or “he can’t walk”). Additionally, we played all of the robot’s speech through headphones instead of a speaker, an approach that is not uncommon when conducting research in classrooms (e.g. see Gordon et al. [69], Hyun et al. [81], and Kory-Westlund et al. [112]). We tried to lessen the potential effects of disembodied speech by using lip-synchronisation and by explaining beforehand that the robot talked through the headphones. Although we do not know how this may have affected the perceived embodiment, children did seem to consistently ascribe the voice to the robot. They often mentioned that it was the robot who spoke to them, gave instruction, and offered feedback. They also speculated about aspects of the robot’s voice (e.g. “he sounds like a boy/girl” or “his voice is like someone my age”). Some stated explicitly that the robot talked to them through the headphones. Nobody mentioned that they had found this to be strange or unpleasant.

For this study we designed a tool for assessing the mindset of young children. The ‘effort’ construct, which emerged from an exploratory factor analysis, showed low internal consistency and was composed of only two items. It was therefore dropped from further analysis in this study. A future version of the questionnaire should focus on further expanding and validating this construct.

Other mindset assessment tools for young children (e.g. Park et al. [164]) measure mindset as a singular bipolar dimension that ranges between a fixed-oriented and growth-oriented mindset. However, we find indications in the data that mindset may instead be considered as two separate unipolar dimensions that range between a less-fixed to more-fixed mindset and a less-growth to more-growth mindset. For example, a single child may score high on growth mindset and at the same time also score high on fixed mindset. Our questionnaire was designed to capture such nuanced situations.

Although the mindset questionnaire has not been rigorously validated as of yet, we have used it in several prior pilot tests with target users involving a wider range of schools. Results of these pilot tests were not inconsistent with teachers’ expectations of
those children. Therefore, we are fairly confident that the tool is sufficiently sensitive to capture differences between individuals with respect to their fixed and growth mindsets. The relatively high scores on the growth construct in the pretest may suggest that the sample from the target user group may have been skewed towards individuals who already exhibit a strong growth mindset. Although mindset is not an explicit part of the school curriculum, the Montessori and Freinet teaching methods of the schools that participated in this study typically have a focus on promoting learning attitudes associated with a growth mindset, which may explain why participating children scored so well on the pretest. In retrospect, these participants may not have been an accurate representation of the user group who potentially has the most to gain from an intervention such as ours. It might therefore be interesting to repeat this study with participants who initially score lower on growth mindset.

7.9 Conclusion

In this chapter we presented a study that investigated effects of effort-related praise on children's growth mindset. Offering praise is an example of a social act that typically only yields effects after long-term sustained interactions. In a longitudinal, in the wild, unsupervised study children interacted at their own initiative with a fully autonomous system situated in their classroom. The system offered effort-related praise while children worked on the learning task. To measure changes in children's mindset before and after the intervention we constructed a questionnaire which was administered as a pretest and posttest interview. The questionnaire consisted of three constructs related to mindset: 1) growth-related items; 2) fixed-related items; and 3) effort-related items.

We deployed two versions of the CAL system: a system that delivered the praise through headphones only, and an otherwise identical CAL system that was extended with a social robot to deliver the praise. A total of 44 children interacted with two consecutive learning tasks over the course of approximately four months. Additionally, we tested a baseline version of this CAL system with 17 children, over the same time span, where there was no robot and where children received no effort-related praise. Children who worked with the latter version showed no significant change in their mindset.

The main research question that guided this work was: What are the effects on the mindset of children when extending an autonomous CAL system with a social robot to deliver effort-related praise? Limitations and discussion points regarding comparability between conditions and extent of the robot's behaviour repertoire arose from the longitudinal in the wild nature of this study. Overall, results showed significant differences on the growth construct between the No-Robot and Robot conditions. The effort-related praise that was delivered by the social robot had a positive effect on the children's growth mindset, whereas the same praise offered by the otherwise identical regular CAL system did not result in a significant effect.

These results offer an interesting insight for social roboticists and educational psychologists working on creating real-world learning interventions. Firstly, we make a methodological contribution to the field of educational HRI by showing the feasibility of conducting comparative studies in real world, longitudinal, unsupervised settings. Secondly, this study makes an empirical contribution by showing the potential benefits of using a robot to more effectively accomplish the social act of delivering supportive praise to promote a growth mindset.

In the previous chapter we showed that the RECAL system had a meaningful im-
impact in short-term interactions. This longitudinal study suggests that a robot and CAL system may also have a meaningful contribution to the learning process in situations that go beyond such single-session encounters. These results are a promising sign for social roboticists and educational psychologists that learning interventions with robots may have a place in real classroom environments.

Next, we use a different lens through which to look at the data collected during this study, to uncover how children worked in practice with the various versions of the CAL system over an extended period of time.
A few months have passed since Chucky joined the class. Thinking back to his first days, he can clearly see how far he has come. He’s managed to help kids progress through thousands of assignments, making some friends along the way.

Something is still nagging him, though. His system has filled up with old logs from previous sessions, but he just cannot bring himself to delete them.

“I wonder,” he mumbles, “perhaps I should take a closer look at all this data, there might be something useful in here.”

8.1 Introduction

This chapter addresses the need for long-term HRI studies. Using the same data collection as described in the previous chapter, we now turn our focus to describing how children worked with the RECAL system during the deployment in their classroom. Embedded sensors in the learning materials allowed the system to automatically guide the learning process, while simultaneously recording children’s interactions with the system. Through these logbooks we show how their learning and their sustained interactions with the system developed throughout the four months, as increasingly challenging learning content unfolded. We examine the extent to which the children could self-regulate this process. Finally, we also share our approach to addressing specific challenges that arose in practice during the study.
In the near future we envision that interactive learning technologies will increasingly find their way into real classrooms. To support long-term interactions they should be capable of offering learning tasks that move and grow with the development of the child and therefore can be accessed across recurring interactions, while the robot behaviour should stay relevant and engaging over time [33]. To understand more about the exact design parameters it is important to gain insight into how children interact with these systems, day-to-day in their own classrooms. However, longitudinal in the wild studies with educational robots are relatively scarce in comparison to the large number of short term studies. A recent review on educational robots found 101 papers with 309 study results [11], but a survey on only longitudinal interactions [128] reported a mere 8 studies. In this chapter we discuss an additional 7 studies, and in a review on evaluation methods for Child Robot Interaction (CRI) we also observed a substantial lack of long-term studies, among other things related to practical challenges in long-term classroom research and the technological readiness of robotic systems [33].

The paucity of long-term studies is a real problem. It is not possible to say whether—and how exactly—insights about CRI transfer from the lab to the real world and to what extent lab studies have succeeded in capturing the richness of interactions as they would develop over time [92].

To advance the field of educational HRI, and move from lab-based single sessions towards real world everyday interactions more studies are needed on how real users interact with robots and learning materials; unsupervised, in the wild, and over extended periods of time. To keep the interest of the children, these type of studies require learning tasks and robot interaction content that are clear and compact enough for using them in short episodes of learning, but at the same time adaptable enough to allow repeated sessions over a longer period. Yet, with the current state of the art in HRI, we do not yet know well enough whether it is feasible to run long-term HRI experiments in the classroom on the basis of a realistic behaviour and task repertoire. We take a first step towards showing the feasibility and potency of running such studies.

8.2 Background and related work

There can be significant differences in the way a user interacts with a robot for the first time, or after a habituation period [128]. Habituation effects are already seen in very young children when presented with novel stimuli [80]. Tanaka et al. [201, 204, 202] have shown that after a familiarisation period, children will have higher quality interactions with a robot as they become accustomed to its behaviour, features and shortcomings; a recognisable routine could play a role here [201]. In order to maintain a child’s interest and engagement the robot should exhibit individual customisation or adaptability [97, 96, 5, 128]. Yet, others reported that 24 months after a disability unit at a school introduced robots in their curriculum people still reported generally positive about the Paro robot, which has no adaptability whatsoever [192].

Different contexts pose unique constraints or challenges that require particular attention to the design of the robot. In a comprehensive survey of 24 papers on long-term studies, Leite et al. [128] found that real-world long-term HRI commonly take place in work environments and public spaces, in health care and therapy, in users’ homes, and in educational settings. Users generally accept robots and are willing to interact with them repeatedly over a period of time. The number of sessions in studies varied considerably between 2-180 (\(M = 32\) sessions, \(SD = 45\)); most studies
reported between 5-20 sessions. Sessions were typically between 6-90 minutes long ($M = 38$ minutes, $SD = 30$) and took place over a period of roughly several weeks to several months ($M = 11$ weeks, $SD = 8$). Leite et al. [128] discuss several longitudinal effects reported by the studies, such as strengthened social interactions and improved well-being of elderly; positive effects in autism therapies; and their potential role in behaviour change support systems. Over time, people may anthropomorphise the robot; they tend to form mental models regarding its behaviour and they may change their preferences regarding interaction and usage patterns. However, the robot did not always live up to users’ initial expectations and during longer studies the importance of regular maintenance became increasingly evident. Finally, Leite et al. [128] identified guidelines for long-term studies regarding the robot’s appearance, the continuity and incremental novelty of its behaviours, its empathy and affective interactions, and its memory and adaptation.

In addition to the application domains outlined above, robots are also being deployed in educational settings. Such robots are often used to investigate how they can contribute to children’s learning. The pioneering work by Kanda et al. [97] investigated how Japanese children interacted with an English speaking robot for two weeks. Their robot could personally identify individual children and would address them by name during the interaction. Their results showed that children often interacted with the robot in groups and that most learning seemed to take place in the second week. However, their results showed a sharp decrease in the frequency of interactions after the first week, suggesting that the robot may not have lived up to children’s initial expectations. Similar effects were found by Leite et al. [129] in an exploratory study on children playing chess with an iCat robot. As the initial novelty wore off, children gave the robot lower social presence ratings and looked at it less frequently. Building on these initial insights, researchers explored ways in which the interaction could be sustained beyond the novelty effects.

In a follow-up study Leite et al. [127] extended their chess-playing iCat robot with a memory of past interactions and an empathic model, which accounted for the child’s affective state and offered supportive feedback and encouragement. The empathic robot seemed to be better at maintaining the interaction over multiple sessions, being able to sustain high levels of perceived presence, engagement, helpfulness and children’s self-validation throughout five interactions. Additionally, the robot’s behaviours aimed at offering esteem and emotional support were preferred most by the children. Similarly, building on their previous work, Kanda et al. [96] investigated whether they could retain children’s interest by extending their robot with capabilities aimed at building social relationships. In particular, their robot incrementally expanded its behaviour repertoire and confided personal matters to individual children who interacted with it more frequently. Their results suggest that children who saw the robot as a peer interacted with it for the full two-month period, forming a friendly relationship, while children who saw it more as a tool became bored after about five weeks.

More recent work has continued to explore additional ways in which to keep children engaged during repeated interactions. For instance, Coninx et al. [43] describe a robot system that supports switching between multiple activities within single sessions, while maintaining a consistent personalised behaviour profile. Children who participated in their case study remained engaged throughout multiple sessions and frequently made use of the opportunity to switch between the three available activities on their own initiative. Their results suggest that such a multi-activity
approach may be a promising way to sustain longer interactions.

A different approach for sustaining motivation during long-term interactions was taken by Jacq et al. [83], who explored the use of a teachable robot to help improve children’s handwriting. Their robot could refine its own handwriting by getting demonstrations and corrections from the child and in doing so the child also improved their own handwriting. Results from individual case studies suggest that the children successfully took on this teaching role and built an affective bond with the robot. The children were committed and engaged with this teaching process and were motivated to focus on the task and pay attention throughout multiple lengthy sessions. Furthermore, Chandra et al. [31] found that children’s perceptions of a robot’s ability to learn is formed over time; only after several sessions were children able to recognise whether their tutoring had actually helped improve the robot’s handwriting or not. Their results suggest that these perceptions may affect the child’s own learning: children showed more improvements in their handwriting when tutoring a robot actually capable of learning.

In contrast with teachable robots, robots can be placed in the role of tutor to more directly instruct the child in one-on-one interactions. Serholt and Barendregt [188] investigated how children’s engagement with a tutoring robot developed over time, looking at moments in the interaction where it initiated a social cue that was outside the scope of the task. Overall, children seemed to be socially engaged with the robot during such cues. The robot consistently elicited gazes, facial expressions, and verbal responses when greeting or asking questions; indications were found that children were affected by the robot’s praise. However, these engagement patterns seemed to decrease somewhat over the course of several interactions. For some children this may have been caused by disappointment over the robot’s limited interaction capabilities.

Tutor robots have also been used during longitudinal interactions in the context of second language learning. Gordon et al. [69] describe a platform aimed at personalising the robot’s tutoring behaviours to the affective state of the child. Overall, children learned new foreign words after several interactions. Furthermore, during the course of their interactions children responded with more positive valence when working with the robot that offered personalised responses. Similarly, Vogt et al. [210] found that children managed to learn foreign words when receiving tutoring from a robot. In their study the robot attempted to further improve its tutoring by using non-verbal iconic gestures (e.g. holding up two fingers when teaching the word “two”), but results showed no additional learning effects of including such gestures.

Although not focused primarily on learning effects, some longer term studies with robots involved younger children. Tanaka et al. [202] looked at social interactions between toddlers and a humanoid robot. They show that, over time, children developed care-taking behaviours towards the robot, progressively treating it more as a peer than as a toy. Kozima et al. [117] found that preschool children displayed similar care-giving behaviours towards a robot with a more minimalist design. Children were initially shy around the robot during the first sessions; in subsequent sessions many children continued interacting with the robot throughout the study and would often play with it together in groups.

These papers show beneficial effects of repeated interactions with educational robots, with a large range of usage contexts, robot platforms used, and interactional complexity and diversity offered. However, most studies report that engagement declines considerably over time. To counter this, it is often suggested that the sys-
tem should gradually show novel behaviours or increasingly varied interactions, but Kennedy et al. [100] showed that increased variations in social behaviours could be detrimental to the core learning process. We explored one aspect of this balancing act by investigating how children progress through two different consecutive learning tasks, while working with an interactive system that displays the same consistent set of behaviours throughout. We try to get a first feeling for the quantity and variation of learning content that is required for sustaining engagement for prolonged periods, and thus for the feasibility of performing long-term HRI studies that may look past the novelty effect.

8.3 Methods and approach

We aimed to analyse the emergent interactions in the long-term unsupervised deployment of the RECAL system, and to share lessons learned from this deployment study. The data reported here was collected during the study reported in the previous chapter, which took place over a duration of approximately four months involving the same 61 children from the same three classrooms.

However, in this chapter we view this data through a different lens and take a more exploratory and broader look towards deploying interactive social learning tools in school. These tools can come in many shapes and forms, of which the presence and absence of a robot and praise are merely specific manifestations. When considering how interactive learning tools are used more generally, these specific aspects become less prominent in the analysis. Therefore, we consider the data from the three CAL system incarnations (baseline, praise no robot, praise robot) as merely variants of a data set from a long-term RECAL deployment. Analysing the aggregated data set allows us to offer broader insights into how children's interactions developed as they became accustomed to having such a setup in their classroom and how they self-navigated the available learning content. Additionally, we share experiences and insights from designing, deploying, working with, and analysing such a setup in practice.

Although a comparative study between robot and non-robot conditions was not our aim here, we were still interested in first insights regarding whether the robot was helpful, damaging, or neither to the long-term interaction. In the previous two chapters we have shown that robots in this kind of task setting can provide real value, transforming the way the children learn. But we do not know whether in the longer run the robot is still helpful, or might in fact detract from the experience once the novelty of the robot wears off and disenchantment sets in. A limited interaction repertoire of a robot might even be worse than a limited interaction repertoire of a non-robot learning system. Therefore, we also briefly consider the similarities and differences in usage patterns across the three versions.

8.3.1 Materials and system overview

The technical setup of the RECAL system used in this study consisted of several key components: two tangible learning tasks with embedded Arduinos and sensors, a tablet interface, a robot, and a desktop machine that orchestrated the fully autonomous interaction with the children in the learning task, as shown in Figure 8.1. The robot was the Zeno R25 from Robokin, an approximately 50cm tall humanoid robot with

Robokind: https://www.robokind.com
expressive facial capabilities. Audio was played through headphones, to not disturb other children working in the classroom.

Previous chapters have offered a detailed description of the system and setup, here we highlight the key aspects that enabled us to follow children's interactions over time.

8.3.1.1 Learning Tasks: Navigating the Difficulty Levels

Respectively, the balance and ramp tasks had 8 and 9 levels of difficulty, through which children could progress at their own pace. As they did so, the system offered verbal task-related instructions. In the lower levels, children were given precise preparatory instructions and were asked for their prediction regarding the outcome (e.g. “place a yellow pot on location 1 and a red pot on location 5, which way do you think it will tilt?”). Higher level assignments were more open-ended (e.g. “find all the balls that are faster than a pingpong ball”). The system offered personalised feedback and praise based on the child’s progress, whether or not they gave a correct answer, and at what level they were doing the task.

To offer appropriate feedback, and to gather more insights into the children's self reported perception of the task and of their own performance, the system also regularly asked the children to comment on their own effort and performance (how hard they worked and whether they thought they gave the right answers), the perceived difficulty of the task, and the desired next assignment difficulty, all through in-task multiple-choice questions on the tablet.

The difficulty levels, feedback, praise, and children's subjective self-reporting were designed to facilitate children's self-regulation of their progress through the difficulty levels, making the tasks build up nicely over time to facilitate longer-term use.
8.3.1.2 Modules for Data Gathering

Data was gathered to carry out the task in an unsupervised way. To detect physical manipulations of the learning materials, they were enriched with embedded sensors. The sensors in the balance scale could detect the tilt of the scale, the placement of pots, and whether the supporting blocks had been removed. The sensors in the rolling ball ramp could detect the angle of each slope, when to release the balls after pressing a button, and the time that it took each ball to reach the finish after release. To get direct input from the children, the system used a tablet with Android 7 running a custom educational app. Besides displaying multi-modal task instructions, this app gathered children's answers to exercises, as well as their responses about self reported difficulty and effort (see above). To identify which child was interacting with the setup, children could scan their RFID card. Additionally, for sensing the presence and location of the child (e.g., for robot gaze) we used a Microsoft Kinect One with SceneAnalyzer software [232].

Using these sources of data, the system automatically measured for every child how often, how long, and at which level they worked on the task. Together with the various self reported subjective data regarding difficulty and effort, this delivered a lot of insight in how children navigated the difficulty levels of the tasks.

8.3.1.3 Robust Interaction Management

All devices were connected to a small desktop Windows 10 PC, hidden out of view, running the RECAL system which orchestrated the interaction based on the various inputs described above. The interaction manager (IM) was based on Flipper, a rule-based information state dialogue engine [139, 215]. The states were modelled around the progression of inquiry phases of each assignment, including error states (used when the child did not follow the proper steps), success states (when a phase was finished and the exercise could proceed to the next phase), and states for the subsequent steps within each phase (explanation by the system; actions by the children). A detailed explanation of this IM was discussed in Chapter 6. Once an exercise was finished and the child wanted to perform another one, the IM would select the (difficulty of the) next exercise based on the child's preference.

We designed the IM for robust multi-modal interaction to enable the system to operate autonomously in a real classroom environment. Rather than capturing crucial user input through state-of-the-art perception and reasoning models, which are often susceptible to environmental and human factors, we designed the core interactions around the tablet interface. Whenever the IM expected input from the child it would display multiple-choice buttons or would display instructions for operating the learning materials. When no actions or responses were detected, the system used timers to repeat instructions or offer additional help. Additionally, the system dealt with various contingencies in the interaction (e.g. user walking away mid-task). Core interactions were extended with more elaborate behaviours to further enrich the learning experience. Through progressive enhancement and graceful degradation techniques we designed additional dynamic behaviours, such as reactive and deictic gaze, depending on whether the system's more advanced sensors were available and were returning reliable values. The core system would continue to operate as best as possible in the event of a malfunctioning module, which would then be automatically restarted at an appropriate point in the interaction.
Table 8.1: Participant demographics. Class 1 worked with a baseline system without robot and without praise, Class 2 without robot and with praise, and Class 3 with robot and with praise.

<table>
<thead>
<tr>
<th>Age</th>
<th>Nr. of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Total (girls/boys)</td>
</tr>
<tr>
<td>Freinet Class 1</td>
<td>8.8 (0.73)</td>
</tr>
<tr>
<td>Montessori Class 2</td>
<td>7.1 (0.83)</td>
</tr>
<tr>
<td>Montessori Class 3</td>
<td>7 (0.82)</td>
</tr>
</tbody>
</table>

8.3.1.4 Robot Behaviour Realisation

The IM requested spoken utterances, robot movement, and tablet interface updates, specified in the Behaviour Markup Language (BML)\(^1\) [110]. The robot could display facial expressions (happiness, surprise), lip synchronisation to speech, interactive gaze (to the child, tablet, or learning materials), and life-like behaviours (blinking). The exact behaviour design of the robot was informed by design guidelines emerging from an extensive contextual analysis of inquiry learning tasks with our target user group \(^5\), previously presented in Chapter 4. The generated BML was sent to ASAPRealizer \(^172\), which further orchestrated the scheduling, planning, and execution of the behaviours; the system’s speech was generated using the Fluency\(^2\) text-to-speech (TTS) engine.

8.3.2 Participants

A total of 61 children between 6-10 years old participated in the study, as described in the previous chapter. The demographics are repeated in Table 8.1 for sake of completeness.

8.3.3 Procedure

The technological setups were installed in each classroom outside of school hours. The next morning, a researcher was briefly present to give an introduction of the system to highlight the various components and illustrate how to initiate a session. Children could work individually with the task on their own initiative and without supervision. All children started on level 3 of 8. The system greeted each child personally by name, after which the tablet displayed textual and visual instructions for completing the assignment. Children could press a “read aloud” button to have the system read the instructions out loud. After each completed assignment, the system asked them whether or not they got the correct answer and if they thought the task was “easy”, “ok”, or “hard”. The system then asked the child if they wanted to continue and whether they wanted to do an easier, equally difficult, or harder assignment. It then loaded a next assignment level according to their selected choice. The system would end the session after four completed assignments. The first task was removed by experimenters after approximately 6-7 weeks.

Two classes had a holiday just after the first task; after that, the experimenters installed and introduced the second task in each class. Now, all children started on

\(^1\)Behaviour Markup Language (BML): \url{http://www.mindmakers.org/projects/bml-1-0/wiki}

\(^2\)Fluency text-to-speech engine for Dutch speech: \url{https://www.fluency.nl/tts/}
level 1 of 9. Although the learning content of this task was different than the first, the interaction procedure and system behaviours were identical. After approximately 8-10 weeks the materials were removed and the study was ended.

8.3.4 Measurements

Measurements were collected automatically on all weekdays, as described in Section 8.3.1.2. This included the quantitative information on how the setups were used in practice as well as the children’s self-reported task performance, subjective assignment difficulty, and their preferred next assignment difficulty. The data was analysed to gain insights on how their usage patterns evolved over time, how they progressed through the available assignment levels, and how well they were able to self-regulate this process.

8.4 Longitudinal deployment in practice: Results

Overall usage statistics per task for each class are shown in Table 8.2. The first learning task, in which children explored the moment of force with a balance scale, was in classes for around 6 to 7 weeks. During that time the 61 children completed a total of 1131 assignments during 371 interaction sessions and worked with the system for an average of 6 minutes per session ($SD = 3.03$ minutes). The majority of children interacted 6 times or more, with several outliers of up to 17 interaction sessions ($M = 6$, $SD = 4$). To compare differences between classes we conducted pairwise Wilcoxon rank sum tests with Bonferroni correction ($\alpha = .0083$). All tests were two tailed. We found that children in class 3 (with robot and praise) initiated significantly more sessions ($M = 7.9$, $SD = 4.5$) than children in class 2 (without robot, with praise) ($M = 4.5$, $SD = 4.2$) ($U = 386.00$, $p = .0062$). No significant differences were found between class 1 (without robot and praise) ($M = 6$, $SD = 2.3$) and class 2 ($U = 137.00$, $p = .017$) nor between class 1 and class 3 ($U = 243.00$, $p = .24$). Figure 8.2 shows a breakdown of sessions and unique users per week for all classes combined. The system was used most frequently during the first weeks, with some individual children even interacting multiple times per week. Overall usage generally declined towards the end; yet we saw that a small number of unique children continued to interact with the system each week. For class 2 and class 3 this decline occurred after roughly 3-4 weeks. However, class 1 had a holiday in the fourth week, during which there were no interactions. The system was deployed in this class for one week longer as compensation; after the holiday we saw a somewhat sustained usage towards the end.

The second learning task, in which children used a ramp to discover potential energy and rolling resistance, was in classes for around 8-10 weeks. Children completed a total of 785 assignments during 274 sessions, with an average session length of 6:44 minutes per session ($SD = 3.02$ minutes). The majority of children initiated four or more sessions ($M = 4.5$, $SD = 2.9$). Similar to the first task, we found that children in class 3 initiated significantly more sessions ($M = 5.8$, $SD = 3.3$) than children in class 2 ($M = 3.1$, $SD = 2.6$) ($U = 346.00$, $p = .0047$). Again, no significant differences were found between class 1 ($M = 4.4$, $SD = 1.8$) and class 2 ($U = 93.00$, $p = .01$) nor between class 1 and class 3 ($U = 260.00$, $p = .17$). The second task was deployed in the schools longer than the first task because several disruptive events took place. Most notable was a two-week holiday, which took place in week 2-3 for class 1, week 5-6 for class 2, and week 6-7 for class 3. Furthermore, towards the end of the school
Table 8.2: Number of schooldays (excluding holidays and weekends) and total number of assignments and sessions in each of the classes. A session was considered completed if the child had finished four assignments. A session was considered stopped when a child indicated to the system they did not wish to continue after completing an assignment. A session was considered abandoned if a child walked away in the middle of a assignment.

<table>
<thead>
<tr>
<th>Task 1 - Balance scale</th>
<th>Total schooldays</th>
<th>Total assignments</th>
<th>Total sessions</th>
<th>Completed sessions</th>
<th>Stopped sessions</th>
<th>Abandoned sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>32</td>
<td>375</td>
<td>111</td>
<td>73</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>Class 2</td>
<td>30</td>
<td>311</td>
<td>103</td>
<td>56</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>Class 3</td>
<td>29</td>
<td>445</td>
<td>157</td>
<td>60</td>
<td>87</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task 2 - Ramp</th>
<th>Total assignments</th>
<th>Total sessions</th>
<th>Completed sessions</th>
<th>Stopped sessions</th>
<th>Abandoned sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>27</td>
<td>235</td>
<td>79</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>Class 2</td>
<td>36</td>
<td>192</td>
<td>62</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>Class 3</td>
<td>35</td>
<td>358</td>
<td>133</td>
<td>33</td>
<td>63</td>
</tr>
</tbody>
</table>

Figure 8.2: A breakdown of the total number of sessions and unique visitors that interacted with the system per week for each of the tasks.

year children were often busy studying and taking exams, and finishing other school assignments. Additionally, class 3 was away practising for a musical in the week before their holiday. We found similar weekly interaction patterns as during the first task; it was popular at the beginning and the number of interactions declined towards the end, although we saw no full abandonment during the study. However, it seems that the decline occurred sooner, after around 2-3 weeks. In week 6, 7, and 8 we saw a slight increase after children returned to school from their holidays. In both tasks we found that class 3 had more than double the number of stopped sessions.

The other measurements regarding self-reported measures and self-regulated pacing revealed no significant differences. We discuss these remaining results using the aggregated data from all children.

In their first interaction with task 1, all children started at level 3. Although children occasionally opted for lower levels, they typically progressed to higher levels as they repeatedly interacted with the system. The number of assignments that were completed were not uniformly distributed across the 8 available levels ($M = 141$ assignments per level, $SD = 108$). This suggests that there may have been a ceiling effect. We found that the highest level was reached by 38 children and was, by far, used most frequently; around 20% of all assignments were completed at this level. Closer investigation reveals that this ceiling effect probably surfaced at around 4-6 interaction sessions, as shown in Figure 8.3.
In the second task children started at level 1, after which they progressed at their own pace in subsequent interactions. In contrast to the first task, the number of assignments completed at each level was more uniformly distributed (M = 87 assignments per level, SD = 16). This suggests that the ceiling effect was less pronounced and seemed to occur later after around 6-7 sessions, as shown in Figure 8.3. This may be due to the fewer number of repeated interactions that took place. Many children didn’t interact frequently enough to exhaust all available levels; only 16 children reached the highest level. It may also be due to the more open nature of the higher levels, where children could repeat the experiment many times in different combinations before arriving at a correct answer.

Regarding self-reported measures, we found no significant differences between the two tasks or the levels. Overall, children reported 73% of their answers to be correct. We found no significant differences in children’s self-reported task performance for any of the levels. Similarly, we found no significant differences between the various levels regarding self-reported subjective difficulty, although the higher levels tended to be rated as “hard” slightly more often than the lower levels. However, we found that children seemed to rate assignments that they had correct as being easier. In those cases, they rated 66% as “easy”, 22% as “ok”, and 12% as “hard”. Assignments they got wrong, on the other hand, were rated 22% as “easy”, 48% as “ok”, and 30% as “hard”. Furthermore, after getting an assignment wrong, children seemed to be more inclined to select a lower level or stick to the same level for their next assignment (39%), compared to when they got the assignment right (26%). These results suggest that the consecutive levels of the tasks were appropriately challenging and that children were able to effectively regulate their own progression.

8.5 Discussion

The results of the four-month study offer insights into how children’s interactions with the learning tasks evolved over time. During the first weeks we found indications of a novelty effect, characterised by a high frequency of use by many different children. This was corroborated during interviews; children mentioned that there was often a queue of classmates who wanted to work with the system. Such novelty effects have long been reported in similar studies, tapering off after a few weeks (e.g., Kanda et al. [97], Fernaeus et al. [63], and Sung et al. [199]). De Graaf et al. [70] showed that
During the novelty phase, users rated their robot less favourably than after they gained some experience with it. They argue that this could be due to the complexity and unfamiliarity of such a novel interaction, which negatively influences the perceived joy and likeability of the interaction. Serholt [187] found similar indications that an overload of information during initial sessions may result in breakdowns in the interaction. To minimise complexity, our learning tasks were carefully designed to be familiar and accessible for children in the target group. When interviewing the children and analysing the results, we found no indications that they encountered particular issues in understanding the system during initial or subsequent interactions.

Furthermore, we found indications for a second novelty effect directly after introducing the second learning task. These results are in line with related work showing how having multiple available activities (e.g. [43]) may help to keep children engaged. In our study, although the core interaction with the system, and especially the dialogue with the robot, remained the same as before, children likely wanted to explore the novel learning content. Other research has shown the merits of gradually introducing novel robot's capabilities (e.g. [96]); yet in our study, we saw sustained, recurring, voluntary interactions triggered by novel learning content without the introduction of novel social and behavioural content.

To ensure availability of a wide variety of learning content and accommodate children's development over time, we implemented several levels in each task. Results show that children were able to successfully self-regulate the pace at which they progressed through the various levels. However, they progressed somewhat quicker than we initially anticipated; after several sessions, many children already reached the highest level. Surprisingly, in the first task, many children continued interacting with the system for several more sessions after reaching that point. Although they could repeat the highest level indefinitely, there was only a limited set of unique assignments available. This may be an indication that these children were still engaged with the system as a whole, even though it offered them no novel content at that point. However, during the second task, we found that many children didn't interact with the system often enough to reach the highest level, and more frequently walked away and abandoned an interaction during a session. This might indicate that they preferred the first task, or had become bored with the interactive system by then and that even the prospect of novel unexplored learning content was not sufficient to keep them engaged. Still, we saw no complete abandonment even in the second task.

We found many similarities in usage patterns across the three versions; children progressed at a similar pace, rated the levels similarly, and reported similar performance. This suggests that these factors may not necessarily be greatly influenced (in either direction) by merely having a robot present. Although we found overall differences in the number of initiated and stopped sessions between classes, we cannot say whether this was caused by the system itself or whether this was a result of innate class dynamics. Differences in class schedules and lesson plans may have played a role, and individual teachers likely imposed different routines and guidelines for when children could work with the system.

8.5.1 Content and Variation: Was it Enough?

Deploying a robot in a classroom for an extended period requires sufficient variation in interactional content [128]. To maintain engagement as children develop their understanding of the educational context, tasks may offer richness, depth, breadth, and
variation. Although implementing such content is not trivial, it appeared manageable and worthwhile. Interest in the system naturally declined over time and resurged after introducing new learning content, but the system continued to attract unique recurring visitors until the end. With this unsupervised deployment study we have gained a first intuition about the extent of the variation that is needed to support a long-term interaction. Our two tasks with each 8-9 levels that adapted to the children’s progress, guided by an interactive system with a limited set of speech phrases and behaviour variations, were sufficient to keep them engaged for multiple interactions during roughly four months on a voluntary basis. We feel this is an order of magnitude that is a good starting point for studying longitudinal educational interactions that extend beyond initial novelty effects. From our standpoint, it is worthwhile investing the time and effort to implement and test such a system and we definitely recommend that more research is carried out on this scale to gather more insights into longitudinal interactions.

8.5.2 Practical Considerations

The system was set in a familiar environment for the children: their own classroom during regular school hours. To make this possible we closely collaborated with the school management and the teachers involved. Although teachers were interested in having the learning task in their class, they expressed not having time to operate it, oversee and coordinate its use, nor help children interact with it. During the course of the study, however, teachers would occasionally get involved with the system. For example, they would make passing references to it during their lessons or they would regulate when children were allowed to use it (e.g. covering it up with a blanket when children were supposed to work on something else). Another important point raised by teachers was that any child interacting with the system should not interfere with regular lessons. The system addressed this by using headphones to play audio. However, especially in the first weeks we found that there were often several bystanders present. Additionally, we needed to be flexible when going to the school to accommodate the classes’ schedules. In some cases this meant that maintenance was delayed and that interviews were spread out over several days or performed at moments when children were not busy with regular school tasks.

Differences between individual teacher’s lesson plans will likely have played a role in the frequency of use. When talking with teachers after the study, all three indicated that they had not included the system in regular lesson schedules. However, teachers had on occasion reminded children that they could play with it at certain times or had prevented them from using it at other times. One teacher had encouraged children to play with it just before the experiment ended, reminding them that it may be their last chance.

Communication with parents was a very important aspect of the study preparation. We organised an information evening to gather their input and discuss any concerns. Feedback was used to improve the information leaflet and consent form. Although most parents were enthusiastic, some raised concerns regarding the educational value. This was resolved in discussions with school management, who agreed that the learning tasks fit as parts within the regular curriculum.

Generally speaking, stakeholders often have similar requirements or constraints, or (unrealistic) expectations regarding the use and deployment of technology. Additionally, users have habits and schedules that must to be taken into consideration when designing the interaction. These issues can be uncovered and addressed at various
moments, for example, by organising information events, focus group sessions, or co-design workshops for users and stakeholders.

8.5.3 Maintenance

Once a system is deployed for an extended amount of time, regular maintenance will likely be necessary [63, 200]. This may involve charging, cleaning and tidying, correcting (technical) failures, or replacing broken or stolen hardware. During this study, most time-consuming maintenance was related to hardware failures (e.g., replacing burnt-out motors in the robot and faulty sensors in the learning task). In some cases the children and teacher invented a story to explain away such events. For instance, at one point the robot’s eyelids had become stuck. Later, children mentioned that the robot was squinting because it had gotten an eye infection. At other times, the robot’s motor sounds would interrupt the teacher. Some children mentioned that it was looking for attention or said it was being rude and rebellious. Other common failures were related to disconnected headphones, USB cables, and power cables, caused by cleaning personnel who would occasionally unplug and move components of the system and forget to plug them back in.

Although interactions took place without direct supervision, they did happen in the classroom where a teacher was present. Likely thanks to this setting we did not encounter instances of children seriously abusing the robot, as can sometimes be the case with in the wild studies (e.g. [157, 225]). There were, unfortunately, instances of missing equipment; often relatively harmless, like children taking marbles from the ramp task. Other times less so. On one occasion someone entered the school and stole one of the tablets. This event made quite an impact on the children and was often mentioned during interviews. Afterwards, the teacher took care of locking away valuable equipment at night.

In this study the researchers doing the maintenance were also involved in recruitment and interviews. Their involvement had therefore to remain concealed so as not to bias children’s responses. To address this we scheduled setup and maintenance at moments after school time with no children present. As a consequence, however, unrecoverable technical errors remained unresolved till the next maintenance window. In general, we recommend that other researchers should think about such auxiliary planning issues and consider the trade offs that are acceptable in their specific study design.

8.6 Conclusion

This chapter addressed the need for more long-term studies in HRI. We identified several challenges, including the robustness of the technical system and interaction design, problems of logistics and organisation, novelty effects, the need for having a sufficient yet feasible amount of learning material to accommodate children’s development over time, and the need for sufficiently varied, but not unnecessarily distractive dialogue content in the child-robot interactions.

By developing a technical setup and deploying it long-term, unsupervised, in the wild, we investigated the feasibility of such studies as well as explore some parameters pertaining to these. In this study, children interacted with the robot and embodied educational materials in the familiar context of their classroom while the system guided them through two consecutive inquiry learning tasks; we analysed their usage patterns over time and compared between variants of the system (with a robot, and without).
Our platform of a robot with limited interaction variation, a tablet interface, and a modest selection of assignments with several sensorised learning materials, seems to be a good starting point for further research. We have shown that this setup is sufficient to sustain recurring, unsupervised and voluntary interactions over extended periods of time (four months). We also showed some limits to this, not completely getting rid of novelty effects, but clearly carrying the study far beyond the typical duration expected for initial novelty effects. There is a clear tapering off of the children’s interest in staying involved, even if many kids stuck around till the end. This provides a starting point for future long-term HRI studies.

Furthermore, it is feasible to automatically gather objective as well as self-report subjective data in an unsupervised way, at sufficient scale to follow individual children over time and run long-term comparative experimental studies. Together with an occasional interview this fine-grained task-by-task data can offer rich insights into the development of children’s interactions with an HRI system.

We showed that, supported by the RECAL system, children could effectively self-navigate the available difficulty levels, consistently progressing from easier to harder assignments. This offers a starting point for more nuanced and personalised guidance in these type of systems.

Finally, if we want to show the real potential of robots in daily life, we need to be able to do long-term comparative studies in which we compare the robot to a non-robot condition, in relatively comparable settings. This is not trivial, but we showed that it is feasible to pursue this: with a still fairly limited robot, the robot and non-robot variants performed somewhat comparably and both managed to “carry children along till the end of the study”. Knowing that, we can now start working towards other long-term studies where we attempt to tease out the actual benefits of (aspects of) robots. This need not necessarily be the same as in short term and/or lab studies – so we need to start carrying out more long-term in the wild studies to start exploring this next level of interesting questions, as has been so well argued by Jung and Hinds [92].
Discussion and Conclusion

“So tell me Chucky,” Mr Donovan leans back in his chair, “what’s the latest report on the students? How has their progress been this year?”

“Ah, I’ve seen many of them improve quite nicely!” Chucky’s eyes light up, “Mees in particular has really turned his game around. I remember he used to get upset if a task didn’t go his way, now there’s no stopping him!” “Oh, speak of the devil, hello Mees! And who is this?”

“This is my cousin, Yara, she’s just visiting during the holidays though.” “So what do you think Yara?” he says proudly.

“Well, I guess he looks cute,” she hesitates, “but you said he was cool..?”

“Ha! We are cool!” he wraps an arm around Chucky’s shoulder, “Welcome to downtown Coolsville. Population: us!”

She laughs, “Well if you say so!” “Do you want to join us for a game of football Chucky?”

“I’m sorry pal,” he replies, “I’m afraid I can’t do that.” “But hit me up if you have some learning to do. For now, the two of you go enjoy your summer break and I’ll see you again next school year, Mees!”

9.1 The birth of the RECAL system

Through the work presented in this thesis we have shown how a carefully designed robot can make a meaningful contribution to children's education. Here, we briefly reflect on the journey towards achieving this goal, through the development of the Robot-Extended Computer Assisted Learning (RECAL) system, as outlined in Figure 9.1.

Making a good interactive system for children is no easy task and there are many things to consider regarding design, study setup, and evaluation at all stages of the product development lifecycle [141]. Before designing any sort of child-robot interaction we first set out to gain a clearer understanding of the context in which our robot would operate. Through ethnographic methods, an exploratory contextual analysis
revealed that children displayed a rich variety of social behaviours on educational, collaborative, and relational levels when working together on an inquiry learning task. We found that the inquiry learning process seemed suitable for eliciting the types of interactions we were interested in exploring with a robot. We proposed concrete design guidelines for a robot to navigate social aspects related to these learning situations. We found social acts in learning to be a promising area where a robot may have an impact.

A research platform was constructed, which would form the basis of the child-robot interaction studies. The platform was geared towards supporting longitudinal in the wild interactions and the inquiry learning tasks were further expanded to fit a wider range of children’s development. These tasks were included in a modular technical architecture that allowed us to iteratively develop and deploy variants of the system ranging in complexity and autonomy. Furthermore, the platform incorporated a collection of measurement instruments, based on observational methods, semi-structured interviews, pictorial tasks, self-reports, and logbooks.

With this research platform we conducted short-term and long-term confirmatory experiments that involved social acts of the robot. Results showed that the robot succeeded in eliciting elaborate and deep explanations from children. Furthermore, results suggested that effort-related praise from the robot could have a positive effect on children’s growth mindset. Specifically, Chapters 6 and 7 show that the robot has made a meaningful impact in these situations; by virtue of being a robot, it contributed to the child’s learning process in a way that a regular interactive learning system could not—through social interaction.

To gain a better understanding of how children worked with the system during
real-world deployments we again turned to exploratory study methods. By putting the system in the hands of real users we uncovered how usage patterns evolved throughout four months of interaction. Conducting a study on such a scale was no simple task, but it was certainly manageable and worthwhile. By doing so we have gained a first feeling for the amount of variation and learning content that may be required to keep children engaged for longer periods, without artificially prolonging the novelty effects of the robot. Children seemed to benefit from the robot being there, without it necessarily being the robot keeping them engaged.

9.2 Discussion and limitations

Although limitations of the individual studies in this thesis have been reported in their respective chapters, there are some broader perspectives we wish to address. In the line of in-the-wild research presented in this thesis we have consistently considered the robot as a holistic entity, with a wide palette of social features and behaviours at its disposal. These social features make it substantially more suitable to perform certain interactions compared to less-social learning tools, as we have shown in our studies. We argued that to understand the potential impact of such a robot, we need to treat the robot as an atomic entity that uses all available social features to their full extent. As a consequence, our work has not focused on pinpointing exactly how each individual modality and social feature influenced the interaction. Our robot, like many other off-the-shelf robots, comes with a certain physical appearance that affords expectations about its social interaction capabilities. This was also something we often heard from children during interviews: “it has ears and eyes so I guess it can hear and see”. Furthermore, our robot had a humanoid appearance and a relatively realistic face. Given the type of natural educational interactions we pursued, it would be strange for the robot not to behave (somewhat) human-like (i.e. showing coherent facial expressions and gaze behaviours, as well as lifelike movements, such as lip-synchronisation and eye blinking). Thus, while designing the social functions of the robot we were compelled to follow the available capabilities offered by the physical form of the robot; in this case function follows form. Not doing so may have undermined the social experience and would have risked an uncanny valley effect [148].

A different line of research focuses on the opposite, where the robot’s intended social function dictates its appearance and behaviours; in that case form follows function. In doing so, researchers can more easily isolate exact modalities and tweak individual behaviours to investigate their effects on the interaction with users, while avoiding uncanny valley effects. In their work on minimalist robot designs, Zaga et al. have shown how these low-anthropomorphic robotings [229] may use gaze behaviours, movement patterns, and context-free utterances to express complex social meaning [231, 228]. Similarly, low-anthropomorphic robots interacting with young children and children with autism have shown promising results through minimalist movements (e.g. [117, 181]).

This research took place in the context of learning; children worked on a learning task, the robot offered advice and feedback, and interactions took place in school. However, critically, this work was not about the robot’s effect on actual learning outcomes. Rather, we have considered specific meta-learning aspects related to social interactions (Chapter 4), explanatory behaviour (Chapter 6), mindset towards learning (Chapter 7), and the progression through tasks (Chapter 8). All these aspects are
important for learning, yet it remains to be seen how well they translate in practice into measurable gains in children’s knowledge and skill.

The overall user-centred design methodology throughout this thesis relied mostly on observational methods, input from domain experts, and user input gathered indirectly through interviews with children and teachers. Therefore, there is a risk of bias in the implemented designs; we may have been looking too closely through “engineering goggles” at certain aspects of the interaction. It can be difficult abandoning a prototype once there is time and energy invested into developing an idea or piece of technology. This can potentially lead to a confirmation bias where the prototype reflects more the train of thought from the developer, than that it accommodates the needs and wishes of the user. The age-old literary advice to “murder your darlings” may also be relevant in such situations, we have found. In an attempt to minimise bias we conducted frequent small-scale pilots with early low-fidelity prototypes of the system, regularly abandoning unfruitful avenues and starting anew. There is a good possibility that other user-centred design methods, such as co-design, focus groups, technology probes, or case studies may have yielded other insights that could have influenced the design in unforeseen ways.

The long-term experiment was situated in real classrooms during school time and was not supervised by a researcher. This can be a problem, since there is less oversight on what actually happens during the study. In these situations it is typically not possible to control for confounds by randomly assigning participants to conditions within a classroom. To somewhat mitigate these effects in our study we confirmed that the teacher’s curriculum did not include any exercises on mindset and we used pretests and posttests to look at changes in children’s mindset on an individual basis. Furthermore, even though it may be possible to choose highly similar classes and schools like in our case, any effects may just be due to group-level differences. That being said, we found that by choosing a different perspective and study methodology it becomes possible to answer a different type of research questions during such situated studies. For instance, as part of the long-term study we also focused on field-testing the system. We used the technology itself as a measuring device to look at children’s emergent interactions with the system in a real-world setting and attempted to learn from that. Doing so, we borrowed from ethnographic methods to follow the development of children’s usage patterns over time.

9.3 Next steps

This thesis started a line of research that has contributions to the broader field of educational HRI. In the two main experiments we compared how children learned with and without a robot. Results show a benefit of having a robot present in a real world learning context, opening many avenues for prospective research.

We poured the best of our knowledge and skill into the design of the robot’s social behaviours, such that it worked well in our task and context. This was just one path towards creating such a robot, however. The next challenge is to replicate the results and gain a better understanding of how we may use robots to reliably trigger the types of learning gains that we encountered. We should explore the details: how does the presentation of the robot, its physical embodiment [132, 170], its social gaze [231], its appearance, its facial expressions [24] and other modalities, its personality, its life-like behaviour, its background story [169], and so forth, affect children’s learning?
Furthermore, we should explore the generalisability by testing different robots and tasks with a more heterogeneous selection of learners in a variety of contexts and cultures. Doing so may yield a strong theoretical contribution for future educational HRI to build upon.

For the four studies reported in this thesis we used a mixture of exploratory design-oriented approaches and confirmatory experimental methodologies to complete a full cycle of designing, testing, and reflecting upon social learning situations involving a robot, as outlined in Figure 9.1. To conduct the studies we developed interactive learning materials, modular technological solutions, and measurement instruments. In doing so, we have constructed a first-generation research platform that has proven to be up to the task for performing longitudinal studies in the wild. We envision that next iterations of the design cycle can build and expand upon this foundation.

The proposed design guidelines of Chapter 4 have helped guide and shape the design of the RECAL system. Many of the illustrative examples accompanying the design guidelines were based on or inspired by initial observations from child-child interactions. As a first step towards validation of the guidelines we looked superficially at children’s social behaviours while they worked with variants of the robot (Chapter 5). We found anecdotal examples that seemed to align with what we might expect given the specific robot and task that were used. However, a deeper empirical analysis has yet to take place of the behaviours and interactions that children display when working with a social robot specifically designed to afford educational, collaborational, and relational interactions. We also do not know whether and how the design guidelines are generalisable; we need more implementations based on specific guidelines to gather insights on what works or not. Future work should further substantiate, validate, and expand upon the design guidelines by systematically comparing situations of child-child with child-robot learning using other robots, educational tasks, application areas, user groups, contexts, and domains. Doing so will contribute to a better conceptual understanding of the complex interplay that exists between the aspects related to social robot design.

Furthermore, to bring the research platform to the next level it is important that the measurement tools are also validated in more diverse contexts. Although Charisi et al. [35] have made a start towards validating the semi-structured interview method aimed at measuring children's perceptions of a robot’s social competence, the remaining measures still need additional work. It might be valuable to bundle efforts with other researchers working on similar novel tools, such as Kory-Westlund et al. [114] who have done promising work on creating and adapting measurement instruments specifically for young children.

Regarding technical aspects, in related projects we have shown that the architecture is flexible and modular enough to support different robots (e.g. [49, 212]) and interaction contexts (e.g. [38, 138]). Improving the accuracy and reliability of interpreted sensor data may contribute to a deeper level of agency for the robot. For educational robots specifically, this involves the sensorised learning materials. In the current platform the interaction models don’t critically rely on the sensordata for generating explicit system responses and scaffolding, since errors in those areas may result in breakdowns in the interaction [187]. Instead, in most cases the system offered only generic suggestions and feedback when it suspected something was wrong (e.g. “Let’s look at the assignment to make sure everything is prepared correctly”) and the models used self-reported measures from the children to offer personalised praise. However, we felt more comfortable basing the system’s less critical implicit responses,
such as reactive gaze, on these sensordata. Once the reliability of reported sensordata is further improved, the robot’s dialogues may include more explicit references, increasing its agency without risking breakdowns in the interaction. For example, the robot may reference individual objects like marbles and pots in its scaffolding (e.g. “You know that red pot should actually go on pin 2 instead of pin 3”).

Furthermore, the memory and interaction models of the robot need to be further expanded to enable scaffolding that adapts to different stages of development and is personalised to individual children. The memory model currently used is basic; it only stores the user’s name and their progress through levels of difficulty. However, by storing more elaborate information about action sequences and performance in tasks it may be possible to offer more exact scaffolding to target areas where the child is struggling. Additionally, by collecting a larger sample of more detailed data it may become possible to construct (generic or personalised) machine learning models that select appropriate forms of scaffolding for various situations.

Inquiry learning has been the pedagogical paradigm underpinning our work. Although this type of task seemed to lend itself well for social learning interactions, there are many other educational domains where social interactions play a role. All these fields utilise aspects of robots’ social capabilities to enhance the learning of children, showing promising results throughout. For instance, robots are becoming popular tools in language learning [32]; the embodied nature of a robot enables it to support the learning content through iconic gestures [221], whereas the social nature of a robot can be used in storytelling situations (e.g. [115, 113]) where it may help modulate the child’s affective state [69]. Additionally, robots are used to help children with their handwriting; these robots evoke the learning-by-teaching paradigm (e.g. [79, 130]) where they achieve success by posing as a social peer that is capable of learning [31]. Robots have also been shown to impact children’s problem-solving skills (e.g. Charisi et al. [36]) and robot peers have supported children’s self-regulation of medical conditions such as diabetes (e.g. Belpaeme et al. [10] and Coninx et al. [43]), where the robot can be provided with a relatable background story to enable more natural co-learning paradigms. As research in these application areas moves towards longitudinal in the wild studies we will start to get a glimpse of the broad impact that a robot may have on the education of our children.

9.4 Concluding remarks: A glimpse of the future of educational robots

Following the framework of Oulasvirta and Hornbæk [161] and Wobbrock and Kientz [222] we contribute to science in the following ways. Firstly, our conceptual contributions are based on expanding the theory on social learning with technology. The ECR model organises theories on learning in such a way that it can be used as a lens through which to view social interactions, as we have shown in the contextual analysis. Furthermore, the design guidelines are a source for further systematic experimentation. These guidelines, and our first exploration of various robot manifestations thereof, offer an initial step towards a broader theoretical understanding of the design space of educational HRI. Secondly, through developing and reflecting on the artefacts that make up the research platform we contribute better understanding of how technology may be geared towards longitudinal, unsupervised, in the wild deployment. The principles underlying the technical architecture allow it to be used in various application areas and contexts, while the measurement instruments allow us to follow
users’ interactions with such systems. Finally, through our empirical results we offer insights in specific social acts where a robot can be leveraged to support learning. More importantly, we show that a social robot can have a meaningful place in education. Through this we paved the way for many more avenues for prospective research.

Throughout this research we have considered typical situations where a child works on a learning task independently, but may still benefit from some social support. We approach this from the direction technology ⇒ social technology where technology is imbued with select social features that serve a distinct goal, rather than the direction social technology ⇐ human where the technology attempts to replicate or even replace social interaction with other humans. The latter is not our goal; social interaction with others plays such a fundamental role in the education and development of children that this is not something we should aim to replace. Instead, we envision that robots take a role in class similar to that of other technologies like tablets or interactive whiteboards: as tools to support or augment the learning experience.

Looking towards the future through rose-coloured glasses we eagerly anticipate robots becoming more robust and fully featured, with ease of use drastically improved. At that point, we envision schools can expand their educational toolbox by buying into an educational robot platform and ecosystem. Robot functionalities may be created and exchanged through a community or marketplace, while being supported by a commercial party. Through training, teachers can design educational activities around the robot, making use of its strengths in particular areas to integrate it in their lesson plans. Much like an educational app on a tablet or laptop, for children the robot becomes another tool that they can work with: in their day-to-day activities they may take a robot from the cupboard, work on a specific task together for a certain amount of time, then return it for the next child to use. All the while, it follows the development of each individual child, offering a consistent and coherent personalised learning experience while making the child’s progress more accessible and transparent to the teacher.

We are not there yet, but the work presented in this thesis is an important step in that direction. Researching child-robot interactions in the wild is notoriously difficult, but can be very rewarding when done right [177]. We have offered a holistic view of designing, implementing, testing, and evaluating an educational robot in such a context. The primary contribution of this work is showing that a carefully designed robot can indeed add a meaningful layer of social interaction to traditional learning tools—the HRI community is definitely on to something!


152


Daniel Davison is a tinkerer and maker at heart, passionate about designing social technology that makes a meaningful impact. He combines a keen eye for technology with a user-centred design approach to create novel interactions attuned to people’s needs and wishes. He received a BSc degree in Computer Science and a MSc degree in Human Media Interaction at the University of Twente.

Besides the work presented in this thesis, Daniel was involved in several other international EU-funded projects. Taking a cross-disciplinary role, he consolidated viewpoints from collaborators and other stakeholders, driving the development of integrated technical solutions. In DE-ENIGMA he developed tools to support appropriate human-robot interaction paradigms for autistic children. Through this, therapists and teachers could help children practice facial features and expressions in a repeatable and predictable setting. In Council of Coaches he worked on a state of the art multi-agent conversational system, in which multiple virtual characters could engage in rich discussions among themselves and with the user. Although initially geared towards healthy lifestyle coaching for older adults, it has since been made available as the open source Agents United platform, enabling researchers to integrate it with a broad range of contexts and applications.

Daniel is involved in the educational programmes Creative Technology and Interaction Technology where he regularly (co-)supervises BSc and MSc graduation projects. He is also the main lecturer for the master courses Trends in Human-Robot Interaction and Advanced Research Project in Human-Robot Interaction, in which students prepare and conduct a study that addresses a relevant topic in the field of HRI. He currently has a position as lab manager at the University of Twente’s Interaction Lab, where he strives to make cutting edge research and technology more accessible within education.

When not in the lab, Daniel enjoys activities like squash, mountainbiking, skiing, hiking, and camping, or simply relaxing at home with a good movie, series, or (board)game. Most of all, he loves spending time outdoors in nature together with his girlfriend and son.
List of publications

The following publications were a core part of this thesis:


Additionally, Daniel was a (co-)author on the following publications:


<table>
<thead>
<tr>
<th>No.</th>
<th>Author (Institution)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Botond Cseke (RUN)</td>
<td>Variational Algorithms for Bayesian Inference in Latent Gaussian Models</td>
</tr>
<tr>
<td>02</td>
<td>Nick Tinnemeier (UU)</td>
<td>Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language</td>
</tr>
<tr>
<td>03</td>
<td>Jan Martijn van der Werf (TUE)</td>
<td>Compositional Design and Verification of Component-Based Information Systems</td>
</tr>
<tr>
<td>04</td>
<td>Hado van Hasselt (UU)</td>
<td>Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference</td>
</tr>
<tr>
<td>05</td>
<td>Bas van der Raadt (VU)</td>
<td>Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.</td>
</tr>
<tr>
<td>06</td>
<td>Yiwen Wang (TUE)</td>
<td>Semantically-Enhanced Recommendations in Cultural Heritage</td>
</tr>
<tr>
<td>07</td>
<td>Yujia Cao (UT)</td>
<td>Multimodal Information Presentation for High Load Human Computer Interaction</td>
</tr>
<tr>
<td>08</td>
<td>Nieske Vergunst (UU)</td>
<td>BDI-based Generation of Robust Task-Oriented Dialogues</td>
</tr>
<tr>
<td>09</td>
<td>Tim de Jong (OU)</td>
<td>Contextualised Mobile Media for Learning</td>
</tr>
<tr>
<td>10</td>
<td>Bart Bogaert (UvT)</td>
<td>Cloud Content Contention</td>
</tr>
<tr>
<td>11</td>
<td>Dhaval Vyas (UT)</td>
<td>Designing for Awareness: An Experience-focused HCI Perspective</td>
</tr>
<tr>
<td>12</td>
<td>Carmen Bratosin (TUE)</td>
<td>Grid Architecture for Distributed Process Mining</td>
</tr>
<tr>
<td>13</td>
<td>Xiaoyu Mao (UvT)</td>
<td>Airport under Control. Multiagent Scheduling for Airport Ground Handling</td>
</tr>
<tr>
<td>14</td>
<td>Milan Lovric (EUR)</td>
<td>Behavioral Finance and Agent-Based Artificial Markets</td>
</tr>
<tr>
<td>15</td>
<td>Marijn Koolen (UvA)</td>
<td>The Meaning of Structure: the Value of Link Evidence for Information Retrieval</td>
</tr>
<tr>
<td>16</td>
<td>Maarten Schadd (UM)</td>
<td>Selective Search in Games of Different Complexity</td>
</tr>
<tr>
<td>17</td>
<td>Jiyan He (UVA)</td>
<td>Exploring Topic Structure: Coherence, Diversity and Relatedness</td>
</tr>
<tr>
<td>18</td>
<td>Mark Ponsen (UM)</td>
<td>Strategic Decision-Making in complex games</td>
</tr>
<tr>
<td>19</td>
<td>Ellen Rusman (OU)</td>
<td>The Mind's Eye on Personal Profiles</td>
</tr>
<tr>
<td>20</td>
<td>Qing Gu (VU)</td>
<td>Guiding service-oriented software engineering - A view-based approach</td>
</tr>
<tr>
<td>21</td>
<td>Linda Terlouw (TUD)</td>
<td>Modularization and Specification of Service-Oriented Systems</td>
</tr>
<tr>
<td>22</td>
<td>Junte Zhang (UVA)</td>
<td>System Evaluation of Archival Description and Access</td>
</tr>
</tbody>
</table>
23 Wouter Weerkamp (UVA), Finding People and their Utterances in Social Media
24 Herwin van Welbergen (UT), Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior
25 Syed Waqar ul Qounain Jaffry (VU), Analysis and Validation of Models for Trust Dynamics
26 Matthijs Aart Pontier (VU), Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots
27 Aniel Bhulai (VU), Dynamic website optimization through autonomous management of design patterns
28 Rianne Kaptein (UVA), Effective Focused Retrieval by Exploiting Query Context and Document Structure
29 Faisal Kamiran (TUE), Discrimination-aware Classification
30 Egon van den Broek (UT), Affective Signal Processing (ASP): Unraveling the mystery of emotions
31 Ludo Waltman (EUR), Computational and Game-Theoretic Approaches for Modeling Bounded Rationality
32 Nees-Jan van Eck (EUR), Methodological Advances in Bibliometric Mapping of Science
33 Tom van der Weide (UU), Arguing to Motivate Decisions
34 Paolo Turrini (UU), Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations
35 Maaike Harbers (UU), Explaining Agent Behavior in Virtual Training
36 Erik van der Spek (UU), Experiments in serious game design: a cognitive approach
37 Adriana Burlutiu (RUN), Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference
38 Nyree Lemmens (UM), Bee-inspired Distributed Optimization
39 Joost Westra (UU), Organizing Adaptation using Agents in Serious Games
40 Viktor Clerc (VU), Architectural Knowledge Management in Global Software Development
41 Luan Ibraimi (UT), Cryptographically Enforced Distributed Data Access Control
42 Michal Sindlar (UU), Explaining Behavior through Mental State Attribution
43 Henk van der Schuur (UU), Process Improvement through Software Operation Knowledge
44 Boris Reuderink (UT), Robust Brain-Computer Interfaces
45 Herman Stehouwer (UvT), Statistical Language Models for Alternative Sequence Selection
46 Beibei Hu (TUD), Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
47 Azizi Bin Ab Aziz (VU), Exploring Computational Models for Intelligent Support of Persons with Depression
48 Mark Ter Maat (UT), Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
49 Andreea Niculescu (UT), Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012 01 Terry Kakeeto (UvT), Relationship Marketing for SMEs in Uganda
02 Muhammad Umair (VU), Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
03 Adam Vanya (VU), Supporting Architecture Evolution by Mining Software Repositories
04 Jurriaan Souer (UU), Development of Content Management System-based Web Applications
05 Marijn Plomp (UU), Maturing Interorganisational Information Systems
06 Wolfgang Reinhardt (OU), Awareness Support for Knowledge Workers in Research Networks
07 Rianne van Lambalgen (VU), When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions
08 Gerben de Vries (UVA), Kernel Methods for Vessel Trajectories
09 Ricardo Neisse (UT), Trust and Privacy Management Support for Context-Aware Service Platforms
10 David Smits (TUE), Towards a Generic Distributed Adaptive Hypermedia Environment
11 J.C.B. Rantham Prabhakara (TUE), Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
12 Kees van der Sluijs (TUE), Model Driven Design and Data Integration in Semantic Web Information Systems
13 Suleman Shahid (UvT), Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
14 Evgeny Knutov (TUE), Generic Adaptation Framework for Unifying Adaptive Web-based Systems
15 Natalie van der Wal (VU), Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.
16 Fiemke Both (VU), Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
17 Amal Elgammal (UvT), Towards a Comprehensive Framework for Business Process Compliance
18 Eltjo Poort (VU), Improving Solution Architecting Practices
19 Helen Schonenberg (TUE), What’s Next? Operational Support for Business Process Execution
20 Ali Bahramisharif (RUN), Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
21 Roberto Cornacchia (TUD), Querying Sparse Matrices for Information Retrieval
22 Thijs Vis (UvT), Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
23 Christian Muehl (UT), Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
24 Laurens van der Werff (UT), Evaluation of Noisy Transcripts for Spoken Document Retrieval
25 Silja Eckartz (UT), Managing the Business Case Development in Interorganizational IT Projects: A Methodology and its Application
26 Emile de Maat (UVA), Making Sense of Legal Text
27 Hayrettin Gurkok (UT), Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
28 Nancy Pascall (UvT), Engendering Technology Empowering Women
29 Almer Tigelaar (UT), Peer-to-Peer Information Retrieval
30 Alina Pommeranz (TUD), Designing Human-Centered Systems for Reflective Decision Making
<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure</td>
<td>Emily Bagarukayo (RUN)</td>
</tr>
<tr>
<td>Qualitative multi-criteria preference representation and reasoning</td>
<td>Wietske Visser (TUD)</td>
</tr>
<tr>
<td>Coalitions in Cooperation Networks (COCOON)</td>
<td>Rory Sie (OUN)</td>
</tr>
<tr>
<td>Evolutionary analysis in PPI networks and applications</td>
<td>Pavol Jancura (RUN)</td>
</tr>
<tr>
<td>On-line Evolution of Swarm- and Modular Robotics</td>
<td>Evert Haasdijk (VU)</td>
</tr>
<tr>
<td>Analysis and Evaluation of Collaborative Modeling Processes</td>
<td>Denis Ssebugwawo (RUN)</td>
</tr>
<tr>
<td>A Collaboration Process for Enterprise Architecture Creation</td>
<td>Agnes Nakakawa (RUN)</td>
</tr>
<tr>
<td>Parameter Tuning and Scientific Testing in Evolutionary Algorithms</td>
<td>Selmar Smit (VU)</td>
</tr>
<tr>
<td>Risk-aware design of value and coordination networks</td>
<td>Hassan Fatemi (UT)</td>
</tr>
<tr>
<td>Information Access for SMEs in Indonesia</td>
<td>Agus Gunawan (UvT)</td>
</tr>
<tr>
<td>Game Design Patterns for Learning</td>
<td>Sebastian Kelle (OU)</td>
</tr>
<tr>
<td>Reflection Amplifiers in self-regulated Learning</td>
<td>Dominique Verpoorten (OU)</td>
</tr>
<tr>
<td>On Combining Alignment Techniques</td>
<td>Anna Tordai (VU)</td>
</tr>
<tr>
<td>A Model and Language for Business-aware Transactions</td>
<td>Benedikt Kratz (UvT)</td>
</tr>
<tr>
<td>Exploration and Exploitation of Multilingual Data for Statistical Machine Translation</td>
<td>Simon Carter (UVA)</td>
</tr>
<tr>
<td>Mining Social Media: Tracking Content and Predicting Behavior</td>
<td>Manos Tsagkias (UVA)</td>
</tr>
<tr>
<td>Handling Abrupt Changes in Evolving Time-series Data</td>
<td>Jorn Bakker (TUE)</td>
</tr>
<tr>
<td>Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions</td>
<td>Michael Kaisers (UM)</td>
</tr>
<tr>
<td>Ontology driven Enterprise Information Systems Engineering</td>
<td>Steven van Kervel (TUD)</td>
</tr>
<tr>
<td>Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching</td>
<td>Jeroen de Jong (TUD)</td>
</tr>
<tr>
<td>News Analytics for Financial Decision Support</td>
<td>Viorel Milea (EUR)</td>
</tr>
<tr>
<td>MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing</td>
<td>Erietta Liarou (CWI)</td>
</tr>
<tr>
<td>Reasoning with Contexts in Description Logics</td>
<td>Szymon Klarman (VU)</td>
</tr>
<tr>
<td>Coordinating autonomous planning and scheduling</td>
<td>Chetan Yadati (TUD)</td>
</tr>
<tr>
<td>Groupware Requirements Evolutions Patterns</td>
<td>Dulce Pumareja (UT)</td>
</tr>
<tr>
<td>The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience</td>
<td>Romulo Goncalves (CWI)</td>
</tr>
<tr>
<td>Quantifying Individual Player Differences</td>
<td>Giel van Lankveld (UvT)</td>
</tr>
<tr>
<td>Making enemies: cognitive modeling for opponent agents in fighter pilot simulators</td>
<td>Robbert-Jan Merk (VU)</td>
</tr>
<tr>
<td>Metagenomic Data Analysis: Computational Methods and Applications</td>
<td>Fabio Gori (RUN)</td>
</tr>
</tbody>
</table>
10 Jeewanie Jayasinghe Arachchige (UvT), A Unified Modeling Framework for Service Design.
11 Evangelos Pournaras (TUD), Multi-level Reconfigurable Self-organization in Overlay Services
12 Marian Razavian (VU), Knowledge-driven Migration to Services
13 Mohammad Safiri (UT), Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
14 Jafar Tanha (UVA), Ensemble Approaches to Semi-Supervised Learning
15 Daniel Hennes (UM), Multiagent Learning - Dynamic Games and Applications
16 Eric Kok (UU), Exploring the practical benefits of argumentation in multiagent deliberation
17 Koen Kok (VU), The PowerMatcher: Smart Coordination for the Smart Electricity Grid
18 Jeroen Janssens (UvT), Outlier Selection and One-Class Classification
19 Renze Steenhuizen (TUD), Coordinated Multi-Agent Planning and Scheduling
20 Katja Hofmann (UvA), Fast and Reliable Online Learning to Rank for Information Retrieval
21 Sander Wubben (UvT), Text-to-text generation by monolingual machine translation
22 Tom Claassen (RUN), Causal Discovery and Logic
23 Patricio de Alencar Silva (UvT), Value Activity Monitoring
24 Haitham Bou Ammar (UM), Automated Transfer in Reinforcement Learning
25 Agnieszka Anna Latoszek-Berendsen (UM), Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
26 Alireza Zarghami (UT), Architectural Support for Dynamic Homecare Service Provisioning
27 Mohammad Huq (UT), Inference-based Framework Managing Data Provenance
28 Frans van der Sluis (UT), When Complexity becomes Interesting: An Inquiry into the Information eXperience
29 Iwan de Kok (UT), Listening Heads
30 Joyce Nakatumba (TUE), Resource-Aware Business Process Management: Analysis and Support
31 Dinh Khoa Nguyen (UvT), Blueprint Model and Language for Engineering Cloud Applications
32 Kamakshi Rajagopal (OUN), Networking For Learning; The role of Networking in a Lifelong Learner’s Professional Development
33 Qi Gao (TUD), User Modeling and Personalization in the Microblogging Sphere
34 Kien Tjin-Kam-Jet (UT), Distributed Deep Web Search
35 Abdallah El Ali (UvA), Minimal Mobile Human Computer Interaction
36 Than Lam Hoang (TUE), Pattern Mining in Data Streams
37 Dirk Börner (OUN), Ambient Learning Displays
38 Eelco den Heijer (VU), Autonomous Evolutionary Art
39 Joop de Jong (TUD), A Method for Enterprise Ontology based Design of Enterprise Information Systems
40 Pim Nijssen (UM), Monte-Carlo Tree Search for Multi-Player Games
41 Jochem Liem (UVA), Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
42 Léon Planken (TUD), Algorithms for Simple Temporal Reasoning
43 Marc Bron (UVA), Exploration and Contextualization through Interaction and Concepts

2014 01 Nicola Barile (UU), Studies in Learning Monotone Models from Data
02 Fiona Tuliyano (RUN), Combining System Dynamics with a Domain Modelling Method
03 Sergio Raul Duarte Torres (UT), Information Retrieval for Children: Search Behavior and Solutions
04 Hanna Jochemmann-Mannak (UT), Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation
05 Jurriaan van Reijsen (UU), Knowledge Perspectives on Advancing Dynamic Capability
06 Damian Tamburri (VU), Supporting Networked Software Development
07 Arya Adriansyah (TUE), Aligning Observed and Modeled Behavior
08 Samur Araujo (TUD), Data Integration over Distributed and Heterogeneous Data Endpoints
09 Philip Jackson (UvT), Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language
10 Ivan Salvador Razo Zapata (VU), Service Value Networks
11 Janneke van der Zwaan (TUD), An Empathic Virtual Buddy for Social Support
12 Willem van Willigen (VU), Look Ma, No Hands: Aspects of Autonomous Vehicle Control
13 Arlette van Wissen (VU), Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
14 Yangyang Shi (TUD), Language Models With Meta-information
15 Natalya Mogles (VU), Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
16 Krystyna Milian (VU), Supporting trial recruitment and design by automatically interpreting eligibility criteria
17 Kathrin Dentler (VU), Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
18 Mattijs Ghijsen (UVA), Methods and Models for the Design and Study of Dynamic Agent Organizations
19 Vinicius Ramos (TUE), Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
20 Mena Habib (UT), Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
21 Kassidy Clark (TUD), Negotiation and Monitoring in Open Environments
22 Marieke Peeters (UU), Personalized Educational Games - Developing agent-supported scenario-based training
23 Eleftherios Sidirourgos (UvA/CWI), Space Efficient Indexes for the Big Data Era
24 Davide Ceolin (VU), Trusting Semi-structured Web Data
25 Martijn Lappendoollar (RUN), New network models for the analysis of disease interaction
26 Tim Baarslag (TUD), What to Bid and When to Stop
27 Rui Jorge Almeida (EUR), Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
28 Anna Chmielowiec (VU), Decentralized k-Clique Matching
29 Jaap Kabbedijk (UU), Variability in Multi-Tenant Enterprise Software
30 Peter de Cock (UvT), Anticipating Criminal Behaviour
31 Leo van Moergestel (UU), Agent Technology in Agile Multiparallel Manufacturing and Product Support
32 Naser Ayat (UvA), On Entity Resolution in Probabilistic Data
33 Tesfa Tegegne (RUN), Service Discovery in eHealth
34 Christina Manteli (VU), The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.
35 Joost van Ooijen (UU), Cognitive Agents in Virtual Worlds: A Middleware Design Approach
36 Joos Buijs (TUE), Flexible Evolutionary Algorithms for Mining Structured Process Models
37 Maral Dadvar (UT), Experts and Machines United Against Cyberbullying
38 Danny Plass-Oude Bos (UT), Making brain-computer interfaces better: improving usability through post-processing.
39 Jasmina Maric (UvT), Web Communities, Immigration, and Social Capital
40 Walter Omona (RUN), A Framework for Knowledge Management Using ICT in Higher Education
41 Frederic Hogenboom (EUR), Automated Detection of Financial Events in News Text
42 Carsten Eijckhof (CWI/TUD), Contextual Multidimensional Relevance Models
43 Kevin Vlaanderen (UU), Supporting Process Improvement using Method Increments
44 Paulien Meesters (UvT), Intelligent Blauw. Met als ondertitel: Intelligencegestuurde politiezorg in gebiedsgebonden eenheden.
45 Birgit Schmitz (OUN), Mobile Games for Learning: A Pattern-Based Approach
46 Ke Tao (TUD), Social Web Data Analytics: Relevance, Redundancy, Diversity
47 Shangsong Liang (UVA), Fusion and Diversification in Information Retrieval

2015
01 Niels Netten (UvA), Machine Learning for Relevance of Information in Crisis Response
02 Faiza Bukhsh (UvT), Smart auditing: Innovative Compliance Checking in Customs Controls
03 Twan van Laarhoven (RUN), Machine learning for network data
04 Howard Spoelstra (OUN), Collaborations in Open Learning Environments
05 Christoph Bösch (UT), Cryptographically Enforced Search Pattern Hiding
06 Farideh Heidari (TUD), Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes
07 Maria-Hendrike Peetz (UvA), Time-Aware Online Reputation Analysis
08 Jie Jiang (TUD), Organizational Compliance: An agent-based model for designing and evaluating organizational interactions
09 Randy Klaassen (UT), HCI Perspectives on Behavior Change Support Systems
10 Henry Hermans (OUN), OpenU: design of an integrated system to support lifelong learning
11 Yongming Luo (TUE), Designing algorithms for big graph datasets: A study of computing bisimulation and joins
12 Julie M. Birkholz (VU), ModiOperandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks
13 Giuseppe Procaccianti (VU), Energy-Efficient Software
14 Bart van Straalen (UT), A cognitive approach to modeling bad news conversations
15 Klaas Andries de Graaf (VU), Ontology-based Software Architecture Documentation
16 Changyun Wei (UT), Cognitive Coordination for Cooperative Multi-Robot Teamwork
17 André van Cleeff (UT), Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
18 Holger Pirk (CWI), Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
19 Bernardo Tabuenca (OUN), Ubiquitous Technology for Lifelong Learners
20 Lois Vanhée (UU), Using Culture and Values to Support Flexible Coordination
21 Sibren Fetter (OUN), Using Peer-Support to Expand and Stabilize Online Learning
22 Zhemin Zhu (UT), Co-occurrence Rate Networks
23 Luit Gazendam (VU), Cataloguer Support in Cultural Heritage
24 Richard Berendsen (UVA), Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
25 Steven Woudenberg (UU), Bayesian Tools for Early Disease Detection
26 Alexander Hogenboom (EUR), Sentiment Analysis of Text Guided by Semantics and Structure
27 Sándor Héman (CWI), Updating compressed column stores
28 Janet Bagorogoza (TiU), Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO
29 Hendrik Baier (UM), Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains
30 Kiavash Bahreini (OU), Real-time Multimodal Emotion Recognition in E-Learning
31 Yakup Koç (TUD), On the robustness of Power Grids
32 Jerome Gard (UL), Corporate Venture Management in SMEs
33 Frederik Schadd (TUD), Ontology Mapping with Auxiliary Resources
34 Victor de Graaf (UT), Gesocial Recommender Systems
35 Jungxiao Xu (TUD), Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction

2016
01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow
03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VU), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UVA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VU), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VU), Trusting Crowdsourced Information on Cultural Artefacts
10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UVA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VU), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UVA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VU), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VU), Refining Statistical Data on the Web
19 Julia Efremova (Tu/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UVA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VU), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UVA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakaratne (VU), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (UvT), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UVA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TUE), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UVA), Monitoring Social Media: Summarization, Classification and Recommendation
36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry
38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UVA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora
43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UVA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis
50 Yan Wang (UVT), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017
01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UVA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VU), Detecting Interesting Differences: Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TUE), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsksy (UVT), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction
16 Aleksandr Chuklin (UVA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UVA), Entity Associations for Search
19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
<table>
<thead>
<tr>
<th>No.</th>
<th>Author (Institution)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Jeroen Linssen (UT)</td>
<td>Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)</td>
</tr>
<tr>
<td>22</td>
<td>Sara Magliacane (VU)</td>
<td>Logics for causal inference under uncertainty</td>
</tr>
<tr>
<td>23</td>
<td>David Graus (UVA)</td>
<td>Entities of Interest — Discovery in Digital Traces</td>
</tr>
<tr>
<td>24</td>
<td>Chang Wang (TUD)</td>
<td>Use of Affordances for Efficient Robot Learning</td>
</tr>
<tr>
<td>25</td>
<td>Veruska Zamborlini (VU)</td>
<td>Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search</td>
</tr>
<tr>
<td>26</td>
<td>Merel Jung (UT)</td>
<td>Socially intelligent robots that understand and respond to human touch</td>
</tr>
<tr>
<td>27</td>
<td>Michiel Joosse (UT)</td>
<td>Investigating Positioning and Gaze Behaviors of Social Robots: People’s Preferences, Perceptions and Behaviors</td>
</tr>
<tr>
<td>28</td>
<td>John Klein (VU)</td>
<td>Architecture Practices for Complex Contexts</td>
</tr>
<tr>
<td>29</td>
<td>Adel Alhuraibi (UvT)</td>
<td>From IT-Business Strategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT</td>
</tr>
<tr>
<td>30</td>
<td>Wilma Latuny (UvT)</td>
<td>The Power of Facial Expressions</td>
</tr>
<tr>
<td>31</td>
<td>Ben Ruijl (UL)</td>
<td>Advances in computational methods for QFT calculations</td>
</tr>
<tr>
<td>32</td>
<td>Thaer Samar (RUN)</td>
<td>Access to and Retrievability of Content in Web Archives</td>
</tr>
<tr>
<td>33</td>
<td>Brigit van Loggem (OU)</td>
<td>Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity</td>
</tr>
<tr>
<td>34</td>
<td>Maren Scheffel (OU)</td>
<td>The Evaluation Framework for Learning Analytics</td>
</tr>
<tr>
<td>35</td>
<td>Martine de Vos (VU)</td>
<td>Interpreting natural science spreadsheets</td>
</tr>
<tr>
<td>36</td>
<td>Yuanhao Guo (UL)</td>
<td>Shape Analysis for Phenotype Characterisation from High-throughput Imaging</td>
</tr>
<tr>
<td>37</td>
<td>Alejandro Montes Garcia (TUE)</td>
<td>WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy</td>
</tr>
<tr>
<td>38</td>
<td>Alex Kayal (TUD)</td>
<td>Normative Social Applications</td>
</tr>
<tr>
<td>39</td>
<td>Sara Ahmadi (RUN)</td>
<td>Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR</td>
</tr>
<tr>
<td>40</td>
<td>Altaf Hussain Abro (VUA)</td>
<td>Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems</td>
</tr>
<tr>
<td>41</td>
<td>Adnan Manzoor (VUA)</td>
<td>Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle</td>
</tr>
<tr>
<td>42</td>
<td>Elena Sokolova (RUN)</td>
<td>Causal discovery from mixed and missing data with applications on ADHD datasets</td>
</tr>
<tr>
<td>43</td>
<td>Maaik de Boer (RUN)</td>
<td>Semantic Mapping in Video Retrieval</td>
</tr>
<tr>
<td>44</td>
<td>Garm Lucassen (UU)</td>
<td>Understanding User Stories - Computational Linguistics in Agile Requirements Engineering</td>
</tr>
<tr>
<td>45</td>
<td>Bas Testerink (UU)</td>
<td>Decentralized Runtime Norm Enforcement</td>
</tr>
<tr>
<td>46</td>
<td>Jan Schneider (OU)</td>
<td>Sensor-based Learning Support</td>
</tr>
<tr>
<td>47</td>
<td>Jie Yang (TUD)</td>
<td>Crowd Knowledge Creation Acceleration</td>
</tr>
<tr>
<td>48</td>
<td>Angel Suarez (OU)</td>
<td>Collaborative inquiry-based learning</td>
</tr>
<tr>
<td>2018 01</td>
<td>Han van der Aa (VUA)</td>
<td>Comparing and Aligning Process Representations</td>
</tr>
<tr>
<td>02</td>
<td>Felix Mannhardt (TUE)</td>
<td>Multi-perspective Process Mining</td>
</tr>
<tr>
<td>03</td>
<td>Steven Bosems (UT)</td>
<td>Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction</td>
</tr>
<tr>
<td>No.</td>
<td>Author (Institution)</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------</td>
<td>-------</td>
</tr>
<tr>
<td>04</td>
<td>Jordan Janeiro (TUD)</td>
<td>Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks</td>
</tr>
<tr>
<td>05</td>
<td>Hugo Huurdeman (UVA)</td>
<td>Supporting the Complex Dynamics of the Information Seeking Process</td>
</tr>
<tr>
<td>06</td>
<td>Dan Ionita (UT)</td>
<td>Model-Driven Information Security Risk Assessment of Socio-Technical Systems</td>
</tr>
<tr>
<td>07</td>
<td>Jieting Luo (UU)</td>
<td>A formal account of opportunism in multi-agent systems</td>
</tr>
<tr>
<td>08</td>
<td>Rick Smetsers (RUN)</td>
<td>Advances in Model Learning for Software Systems</td>
</tr>
<tr>
<td>09</td>
<td>Xu Xie (TUD)</td>
<td>Data Assimilation in Discrete Event Simulations</td>
</tr>
<tr>
<td>10</td>
<td>Julienka Mollee (VUA)</td>
<td>Moving forward: supporting physical activity behavior change through intelligent technology</td>
</tr>
<tr>
<td>11</td>
<td>Mahdi Sargolzaei (UVA)</td>
<td>Enabling Framework for Service-oriented Collaborative Networks</td>
</tr>
<tr>
<td>12</td>
<td>Xixi Lu (TUE)</td>
<td>Using behavioral context in process mining</td>
</tr>
<tr>
<td>13</td>
<td>Seyed Amin Tabatabaei (VUA)</td>
<td>Computing a Sustainable Future</td>
</tr>
<tr>
<td>14</td>
<td>Bart Joosten (UVT)</td>
<td>Detecting Social Signals with Spatiotemporal Gabor Filters</td>
</tr>
<tr>
<td>15</td>
<td>Naser Davarzani (UM)</td>
<td>Biomarker discovery in heart failure</td>
</tr>
<tr>
<td>16</td>
<td>Jaebok Kim (UT)</td>
<td>Automatic recognition of engagement and emotion in a group of children</td>
</tr>
<tr>
<td>17</td>
<td>Jianpeng Zhang (TUE)</td>
<td>On Graph Sample Clustering</td>
</tr>
<tr>
<td>18</td>
<td>Henriette Nakad (UL)</td>
<td>De Notaris en Private Rechtspraak</td>
</tr>
<tr>
<td>19</td>
<td>Minh Duc Pham (VUA)</td>
<td>Emergent relational schemas for RDF</td>
</tr>
<tr>
<td>20</td>
<td>Manxia Liu (RUN)</td>
<td>Time and Bayesian Networks</td>
</tr>
<tr>
<td>21</td>
<td>Aad Slootmaker (OUN)</td>
<td>EMERGO: a generic platform for authoring and playing scenario-based serious games</td>
</tr>
<tr>
<td>22</td>
<td>Eric Fernandes de Mello Araujo (VUA)</td>
<td>Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks</td>
</tr>
<tr>
<td>23</td>
<td>Kim Schouten (EUR)</td>
<td>Semantics-driven Aspect-Based Sentiment Analysis</td>
</tr>
<tr>
<td>25</td>
<td>Riste Gligorov (VUA)</td>
<td>Serious Games in Audio-Visual Collections</td>
</tr>
<tr>
<td>26</td>
<td>Roelof Anne Jelle de Vries (UT)</td>
<td>Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology</td>
</tr>
<tr>
<td>27</td>
<td>Maikel Leemans (TUE)</td>
<td>Hierarchical Process Mining for Scalable Software Analysis</td>
</tr>
<tr>
<td>28</td>
<td>Christian Willemse (UT)</td>
<td>Social Touch Technologies: How they feel and how they make you feel</td>
</tr>
<tr>
<td>29</td>
<td>Yu Gu (UVT)</td>
<td>Emotion Recognition from Mandarin Speech</td>
</tr>
<tr>
<td>30</td>
<td>Wouter Beek</td>
<td>The &quot;K&quot; in &quot;semantic web&quot; stands for &quot;knowledge&quot;: scaling semantics to the web</td>
</tr>
</tbody>
</table>

**2019**

01 Rob van Eijk (UL), Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TUE), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TUE), Process Mining with Streaming Data
06 Chris Dijkshoorn (VU), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes
09 Fahimeh Alizadeh Moghaddam (UVA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs
12 Jacqueline Heinerman (VU), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TUE), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN), Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TUE), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VU), Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VU), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VU), Pragmatic factors in (automatic) image description
26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OUN), The Gamification Design Process applied to (Massive) Open Online Courses
28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VU), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VU), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TUE), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TUE), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OUN), Gamification with digital badges in learning programming
36 Kevin Ackermans (OUN), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OUN), Learning visually grounded and multilingual representations

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding
04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TUE), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OUN), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL), Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TUE), In-database Preprocessing for Process Mining
11 Sepideh Mesbahi (TUD), Semantic-Enhanced Training Data Augmentation Methods for Long-Tail Entity Recognition Models
12 Ward van Breda (VU), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OUN), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OUN), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VU), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VU), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nobrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots
25 Xin Du (TUE), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Bibeg Limbu (TUD)</td>
<td>Multimodal interaction for deliberate practice: Training complex skills with augmented reality</td>
</tr>
<tr>
<td>29</td>
<td>Ioan Gabriel Bucur (RUN)</td>
<td>Being Bayesian about Causal Inference</td>
</tr>
<tr>
<td>30</td>
<td>Bob Zadok Blok (UL)</td>
<td>Creatief, Creatieve, Creatiefst</td>
</tr>
<tr>
<td>31</td>
<td>Gongjin Lan (VU)</td>
<td>Learning better – From Baby to Better</td>
</tr>
<tr>
<td>32</td>
<td>Jason Rhuggenaath (TUE)</td>
<td>Revenue management in online markets: pricing and online advertising</td>
</tr>
<tr>
<td>33</td>
<td>Rick Gilsing (TUE)</td>
<td>Supporting service-dominant business model evaluation in the context of business model innovation</td>
</tr>
<tr>
<td>34</td>
<td>Anna Bon (MU)</td>
<td>Intervention or Collaboration? Redesigning Information and Communication Technologies for Development</td>
</tr>
<tr>
<td>35</td>
<td>Siamak Farshidi (UU)</td>
<td>Multi-Criteria Decision-Making in Software Production</td>
</tr>
<tr>
<td>2021 01</td>
<td>Francisco Xavier Dos Santos Fonseca (TUD)</td>
<td>Location-based Games for Social Interaction in Public Space</td>
</tr>
<tr>
<td>02</td>
<td>Rijk Mercuur (TUD)</td>
<td>Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models</td>
</tr>
<tr>
<td>03</td>
<td>Seyyed Hadi Hashemi (UVA)</td>
<td>Modeling Users Interacting with Smart Devices</td>
</tr>
<tr>
<td>04</td>
<td>Ioana Jivet (OU)</td>
<td>The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning</td>
</tr>
<tr>
<td>05</td>
<td>Davide Dell'Anna (UU)</td>
<td>Data-Driven Supervision of Autonomous Systems</td>
</tr>
<tr>
<td>06</td>
<td>Daniel Davison (UT)</td>
<td>&quot;Hey robot, what do you think?&quot; How children learn with a social robot</td>
</tr>
<tr>
<td>07</td>
<td>Armel Lefebvre (UU)</td>
<td>Research data management for open science</td>
</tr>
<tr>
<td>08</td>
<td>Nardie Fanchamps (OU)</td>
<td>The Influence of Sense-Reason-Act Programming on Computational Thinking</td>
</tr>
<tr>
<td>09</td>
<td>Cristina Zaga (UT)</td>
<td>The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play</td>
</tr>
<tr>
<td>10</td>
<td>Quinten Meertens (UvA)</td>
<td>Misclassification Bias in Statistical Learning</td>
</tr>
<tr>
<td>11</td>
<td>Anne van Rossum (UL)</td>
<td>Nonparametric Bayesian Methods in Robotic Vision</td>
</tr>
</tbody>
</table>