Going Through the Motions: Evaluating the Impact of Task, Device and Platform on Mouse-Tracking Derived Measures of Decision Difficulty

by

Alexandra A. Ouellette Zuk

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Neuroscience and Mental Health Institute

University of Alberta

Abstract

As decisions require actions to have an effect on the world (Cisek & Kalaska, 2010), measures derived from movements can be used to provide a powerful index of decision-making processes (e.g., Gallivan & Chapman, 2014). Measures of trajectory curvature (interpreted as a competitive pull from the non-chosen choice; Spivey, Grosjean, & Knoblich, 2005), reaction time, and movement time obtained during mouse-tracked, reach-decision tasks thus provide a metric of the relative difficulty of decisions (McKinstry, Dale, & Spivey, 2008). While these measures of decision difficulty have been demonstrated across a variety of decision domains, they are reported in different studies with different groups of participants and are often captured using experimental systems both impractical and inaccessible outside of laboratory exploration. The current study therefore aimed to assess whether within-participant metrics of decision difficulty remain consistent across decision domains varying in choice stimuli, objectivity and processing requirement, data collection devices varying in size and user-interaction requirements (e.g., mouse-based interactions to touchscreen use) and implementation platforms requiring individualized data processing and cleaning strategies. Specifically, three primary questions were addressed: 1) How do measures of decision difficulty change across testing device: computers, tablets and smartphones? 2) How do measures of decision difficulty relate to each other and how does this change across decision domain and device? and 3) How does implementation platform effect measures of decision difficulty? Deploying a classic mouse-tracking, reach-decision paradigm, participants ($N = 279$) were asked to complete a numeric-size congruency (SC) task requiring objective perceptual judgements of which of two digits with different physical sizes was numerically larger (Faulkenberry *et al.,* 2016), a sentence verification (SV) task requiring semi-subjective conceptual judgements about the truth value of statements varying in truth value

and negation (Maldonado *et al.,* 2019), and a photo preference (PP) task requiring a subjective judgement of preference between two images varying in pleasantness (Koop & Johnson, 2013). An identical experiment was developed for implementation using both Labvanced and Horizon testing platforms, with participation using the prior platform distributed between personal computer ($N = 83$), tablet ($N = 78$) and smartphone ($N = 78$) testing devices and the latter limited to personal computer use ($N = 40$). Participation occurred remotely, online, and without device specification requirements.

Broadly, task-specific results replicated previous work: SC: We found an increase in decision difficulty when digit choice options were incongruent in physical and numeric size; SV: Measures of decision difficulty increased when participants were asked to affirm negated sentences compared to non-negated sentences, with greater negation-driven difficulty effects for true statements than false statements; PP: Images matched in pleasantness showed increased decision difficulty compared to image options that differed in pleasantness. Importantly, task-dependent decision difficulty effects were replicated independent of testing device or platform, demonstrating the robustness of trajectory-tracked measures of decision difficulty and offering seminal validation for the study of decision processes using small, portable devices outside of controlled laboratory spaces. Independent from these replication results, nuanced differences observed in pre-movement (i.e., reaction time) and post-movement (i.e., movement time and trajectory curvature) measures revealed device-dependent differences in which tablet- and smartphone-acquired measures showed right-hand reach direction biases resembling those seen in real-world movements while computer-acquired results did not. Tablet- and smartphone-use also showed greater sensitivity to decision difficulty expressed in movement times and trajectory curvature while computer-acquired results displayed greater sensitivity to decision difficulty expressed in reaction times. Finally, while task-replication results revealed an increase in all measures (reaction time, movement time and trajectory curvature) in response to increased decision difficulty, a correlation analysis between measures of decision difficulty revealed consistent between-measure relationships within each task and across each device and platform, wherein faster decisions (i.e., decisions with decreased reaction time) had more decision difficulty reflected in the movement (increased movement time and trajectory curvature). Together, these results provide support for models of decision making in which decision processes continue to unfold after movements to enact a choice have been initiated (Wispinski, Gallivan & Chapman, 2020), and further suggest that these processes are flexibly adjusted along the time course of a decision even when decision domain and difficulty remain consistent.

Preface

This thesis is an original work by Alexandra A. Ouellette Zuk. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "ACE 2", Pro00087329, 2/25/2019.

Acknowledgements

As I come to the end of my thesis project, I find myself feeling incredibly indebted to all those who have supported me along the way. I would first and foremost like to express my deepest gratitude to Dr. Craig Chapman, whose unwavering encouragement, guidance, and openness has extended his role beyond academic supervisor to true mentor. It has been an amazing experience having you in my corner not only at every stage of this research project but also as I navigated graduate studies, a global pandemic, and all the uncertainties that came with them. I attribute much of both my personal and professional growth over the last few years to the genuine, caring and comfortable atmosphere you have established at the ACELab and the positive example you have set for your students.

I am also sincerely grateful for all my fellow lab mates – Jenn Bertrand, Nathan Wispinski, Ewen Lavoie, Scott Stone and Brea Chouinard – as well as Riley Dawson and Quinn Boser, who have provided advice, technical aid and a lot of laughs. You are a truly exceptional group, and it has been a pleasure sharing a (recently online) lab space, golfing and gold-panning with all of you. May hot-take time never cool down! Additional thanks must also go to Jenn and Nathan for all their patient and readily given help in all things statistics, from theory to analysis script writing. You helped me avoid many a headache during this research venture.

Many thanks to the members of my committee – Dr. Anthony Singhal, Dr. Tanya Berry and Dr. Patrick Pilarski – for your feedback and guidance in producing this thesis, and to Amber Lapointe, the administrative backbone of the NMHI, for always having an enthusiastic answer to my many questions and being so quick with a kind and supportive word. I would also like to extend my sincere thanks to Jamie Betts, who graciously supported my internship at Neurosight and provided direction in the development of this project. It has been a great pleasure working with you and your company, and I am very appreciative of the invaluable exposure to business development and the inner workings of the assessment industry that you provided.

My thanks and appreciation also go to the Neuroscience and Mental Health Institute, the National Sciences and Engineering Research Council of Canada, Mitacs Accelerate International and Neurosight for the funding and resources that made this work possible.

Finally, I am incredibly grateful to my parents, Gail and Denis, my siblings, Joel and Cam, and my partner in all things, Mat, who have been and always will be my biggest supporters. Your sacrifices, unconditional love and unwavering confidence in me has been a true blessing.

Table of Contents

List of Tables

List of Figures

1.0 – Introduction

1.0 - Introduction

Our daily lives unfold as an amalgamation of decisions made and actions taken to execute those choices. Ultimately, these actualized choices shape our lives and our societies. As a result, the pursuit of understanding human decision behaviour has inspired researchers for centuries, from interest in risk preference amongst gamblers (Bernoulli, 1954), to willingness to pay given prior value contexts (Khaw, Glimcher, & Louie, 2017). Historically, measures of decision making have often been derived from verbal reports (as in Khaw *et al.*, 2017), or inferences about observed choices (often involving a limited set of options, e.g., Padoa-Schioppa, & Assad, 2006). More recently, discrete measurements of behavioural outcomes, such as reaction time and accuracy, have been popularized for their ability to generate rich accounts of cognitive processing (see Schulte-Mecklenbeck *et al.*, 2017, for review). Where these off-line measures fall short, however, is in their ability to provide insight into the evolution of cognitive processes over time, or whether the temporal convergence of processes are responsible for driving responses. An expansive understanding of these features can instead be found in analysis of the *dynamics* of behavioural output over time, captured as a continuous stream of non-conscious movement behaviours occurring over the course of a decision process (Freeman, Dale & Farmer, 2011; Gallivan & Chapman, 2014; Wispinski, Gallivan & Chapman, 2020; Gallivan, Chapman, Wolpert & Flanagan, 2018).

1.1 – Towards an integrated view of cognition and movement

Classic theories view the mind as a computational machine in which cognition and action arise from functionally independent systems, with perception informing cognition and cognition informing action in a hierarchical manner (see Rosenbaum, 2005, for discussion). Motor movements were therefore seen as the product of lower-order systems that could provide little information about the higher-order processes that preceded them. This unidirectional, discrete-processes approach to human cognition and action has since been deemed problematic as it cannot account for phenomena such as cognitive tuning, in which motor behaviours can inform cognitive processing (e.g., rating things more positively if facial muscles are in smiling-related positions, Strack, Martin & Stepper, 1988; see Koop & Johnson, 2013, and the references within). Further, recent work demonstrating correlated activity between cognitive and motor brain regions provides support for integrated, rather than discrete, cognitive and motor processes in which movement is continually updated by cognitive processing over time. For example, Freeman, Ambady, Midgley and Holcomb (2011) demonstrated that in tasks involving stimuli categorization, lateralized readiness potentials and N300/N400 event-related potentials occur in parallel independent of whether an action would be required to manually indicate said categorization. The first of these ERPs reflect a preparation for motor activity and the second reflects the accumulation of stimuli characteristics for categorization (e.g typical vs atypical), suggesting that cognitive categorization processes automatically and continuously update the motor cortex to guide hand-movement responses over time.

The intimate relationship between action and cognitive dynamics is further exemplified in studies demonstrating simultaneous representation of multiple grip types in the primate anterior intraparietal area prior to specification of grip required for object interaction (Baumann, Fluet & Scherberge, 2009), and again in the simultaneous representation of potential choice targets in the primate dorsal premotor cortex prior to the correct target being indicated for final action (Cisek & Kalaska, 2005). Together, these findings support an integrated view of cognitive and motor systems, whereby mental processes direct manual dynamics on a continuous scale.

The implication of the integrated nature of these two systems is, then, that simple bodily movements can provide real-time read-outs of concurrent cognitive processing. Tracking eye movement has been used for this purpose for decades, serving as a proxy for tracking attention and informing inferences about thought processes (Yarbus, 1967; Just & Carpenter, 1980). Similarly, changes in hand movements - and by extension, mouse trajectories - have been shown to reflect ongoing mental processes with high temporal sensitivity (Ghez *et al.*, 1997; Song & Nakayama, 2009; Freeman, 2018). Moreover, as decisions require actions in order to have an effect on the world (Cisek & Kalaska, 2010), these motoric measures can be used to provide a powerful index of the decision-making process (e.g., Chapman *et al.*, 2010a-b; Gallivan & Chapman, 2014; Wispinski, Gallivan & Chapman, 2020). In the current study, use of these measures will enable inferences about the decision made (e.g., relative difficulty) as a function of the movement dynamics used to enact a particular choice.

1.2 – Decisions as dynamic competition

Prominent models of decision making (e.g., drift-diffusion and race models; Heath & Link, 1975; Smith & Vickers, 1988; see Wispinski, Gallivan & Chapman, 2020, for a recent review) position decision making as a process during which evidence (information relevant to a decision, reflected within task-dependent neural activity; Platt and Glimcher, 1999; Huk and Shadlen, 2005; Hunt *et al.,* 2018) is accumulated over time until support for one particular option reaches a threshold, at which time a decision favouring that option is made (see Figure 1.1 A). Within these models, the accumulation of evidence towards one option over the other is a dynamic process as our representation of our choices evolve (e.g., Cunningham, Dunfield, & Stillman, 2013). This evolution arises as representational information is integrated from both external and

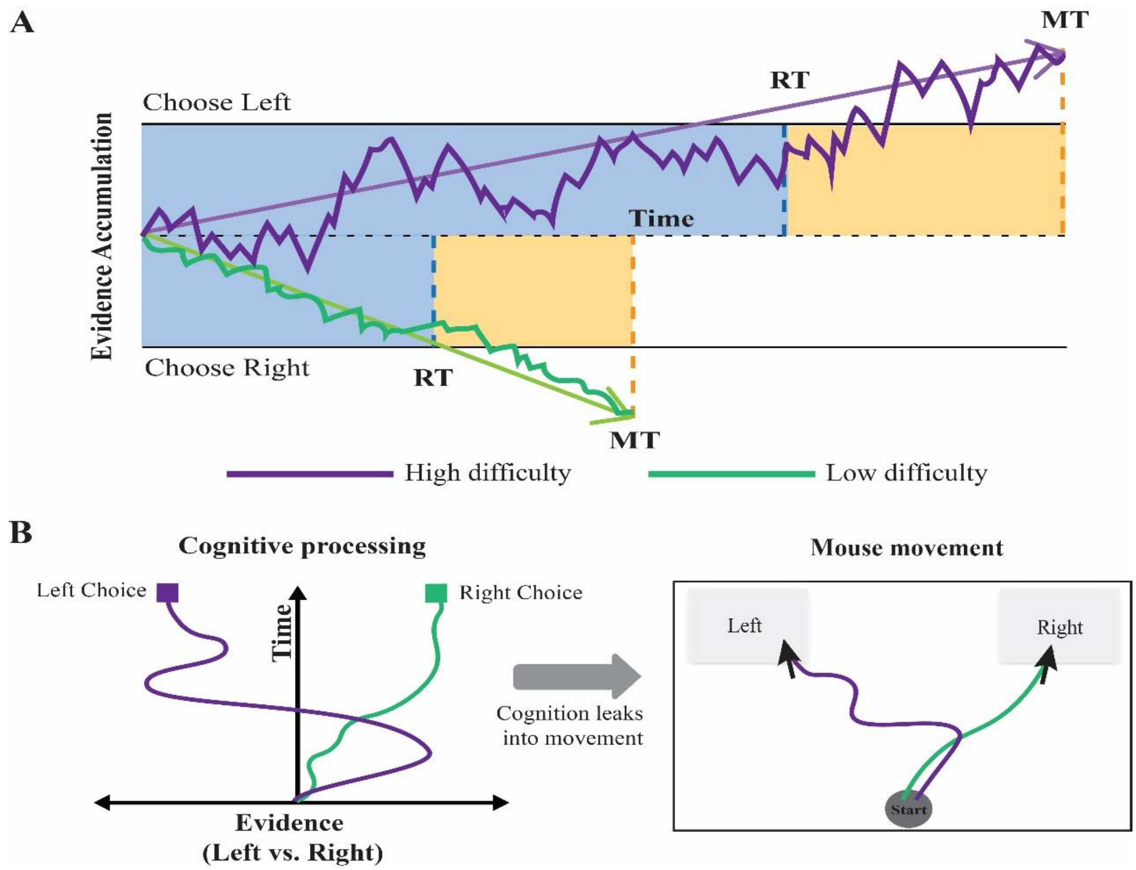

Figure 1.1 A schematic of the relationship between evidence accumulation models of decision making and its physical manifestation in mouse movements. **A)** An evidence accumulation model assumes that, over the course of a decision, relative evidence in support of one option or another (purple and green lines) are noisily accumulated over time until it reaches a predetermined threshold for selection in that direction (black boundary lines). Presented are two examples, one where evidence strongly and consistently supports selection of the right option (Low difficulty, green line) and one where evidence weakly and inconsistently supports selection of the left option (High difficulty; purple line). The average slope of evidence accumulation is represented by the arrows in the corresponding colour. Note that classic evidence accumulation models depict a termination of evidence accumulation at reach onset (i.e., limited to RT; blue, terminating at the dashed blue line) but here we demonstrate a continuation of evidence accumulation beyond the movement threshold into movement time (yellow) until the reach has terminated (MT; dashed yellow line). Adapted from Stillman *et al.,* 2020. **B)** When simplified, the evidence accumulation processes can be depicted as an activation difference between two options as a function of time (on the left). Over the course of a classic reach-decision mouse tracking paradigm, continuous mouse movements (depicted as the recorded mouse cursor position; on the right) reflect the relative activation of choice options, such that decision processes inform cursor trajectories. More difficult decisions, where evidence is less consistent towards one option and the activation difference between options fluctuates demonstrate greater trajectory curvature in the corresponding mouse movement (purple). In contrast, decisions with low difficulty, where there is little fluctuation in activation difference between options, demonstrate relatively direct corresponding mouse trajectories (green). Adapted from Schoemann *et al.*, 2019.

internal sources (e.g., environment sampling vs. memories; Shadlen & Shodamy, 2016), has dimensional weightings applied (e.g., gains vs. losses; Chapman *et al.*, 2015), and multiple levels of choice representation are compared and blended (e.g., good- vs. action- based representations; Cisek, 2012; Chen & Stuphorn, 2015).

Over the dynamic, continuous time course of a decision, should one option provide strong, consistent evidence for its selection when the other does not, this manifests as a strong competitive pull towards one option and thus a relatively easy decision (see green traces in Figure 1.1). However, should the competitive pull of each of the choice options be relatively equal in strength - neither accumulating evidence stronger or more consistently than the other - this manifests as a difficult choice (see purple traces in Figure 1.1). Within this framework, we can therefore define decision difficulty by the strength of the evidence accumulated to bolster selection of one choice option over another (i.e., the relative difference in competitive pull towards each option).

While the accumulation of evidence towards one choice option or another was once thought to terminate with movement onset (i.e. the decision is resolved within the reaction time prior to movements made to enact the decision, consistent with a serial account of cognition and movement; Smith & Vickers, 1988; Ratcliff & Rouder, 1998; Krajbich & Rangel, 2011), recent work instead supports the continued integration of evidence *during* movement (i.e. the decision continues to be resolved as movement unfolds; Ghez *et al.,* 1990; Song and Nakayama, 2009; recent works reviewed by Wispinski, Gallivan and Chapman, 2020).

Importantly, the continued contribution of decision-making during movement becomes apparent when examining the physical movements used to enact a choice, with reaches towards an option chosen for selection serving as a reflection of the evidence accumulation process (Stillman *et al.*, 2020) such that less direct, less consistent movements reflect increased decision difficulty (Schoemann *et al.,* 2019). As such, decisions are dynamic not only in the way they unfold as an internal cognitive process, but also in the way they are expressed through movement.

1.3 – Capturing ongoing decisions through movement

Taking advantage of the reciprocal relationship between cognition and movement, classic trajectory tracking techniques involve the presentation of spatially separated response options and continuous recording of motor trajectories during option selection. Curvature (i.e., deviation from a direct path towards a selected option) in the trajectories of responses are interpreted as a competitive pull towards the non-chosen alternative choice (Spivey, Grosjean, & Knoblich*,* 2005), driven by evidence accumulated during ongoing decision processes. The extent and time course of curvature can then be compared against theoretical accounts that make predictions about processes taking place over the time-course of a mental activity.

This general paradigm has proven valuable in the understanding of a multitude of behaviours from domains such as social cognition (e.g. stereotyping, Freeman & Ambady, 2010; precepts of race, Freeman, Pauker & Sanchez, 2016; and gender biases, Freeman & Johnson, 2016), language processing (e.g. phonetic competition, Spivey *et al.,* 2005; syntactic expectation, Farmer, Cargill, Hindy, Dale, & Spivey, 2007; and speech perception, Spivey *et al*., 2005) and numeric operations (mental arithmetic, Szaszi *et al*., 2018; Pinheiro-Chagas *et al.*, 2017; relative numeric representations, Erb, Moher, Song & Sobel, 2018). Importantly, trajectory tracking research has also revealed new insights into decision making processes themselves, with analyses of trajectories enabling inferences about relative decision difficulty as a function of trajectory curvature (Koop & Johnson, 2013; McKinstry, Dale, & Spivey, 2008; Faulkenberry, 2014). In a study by McKinstry and colleagues (2008), for example, participants indicated whether they agreed ("Yes") or disagreed ("No") with propositions ranging from high truth value (decidedly true, e.g. "Should you brush your teeth every day?"), to medium truth value (ambiguous truth, e.g. "Is murder sometimes justifiable?), to low truth value (decidedly false, e.g. "Is a thousand more than a billion?"). The authors found that questions with low truth values showed significantly more trajectory curvature (measured as maximum deviation) during choice selection than the trajectories of high truth value questions (see Figure 1.2 for results), indicating greater attraction to the "Yes" alternative even while participants responded "No". This was taken to suggest that evaluating a proposition as false is more difficult than evaluating a proposition as true. Studies such as this set the precedent for the use of trajectory tracking to make relative judgements about individual decisions as a function of the degree of curvature exhibited during the reach-decision task.

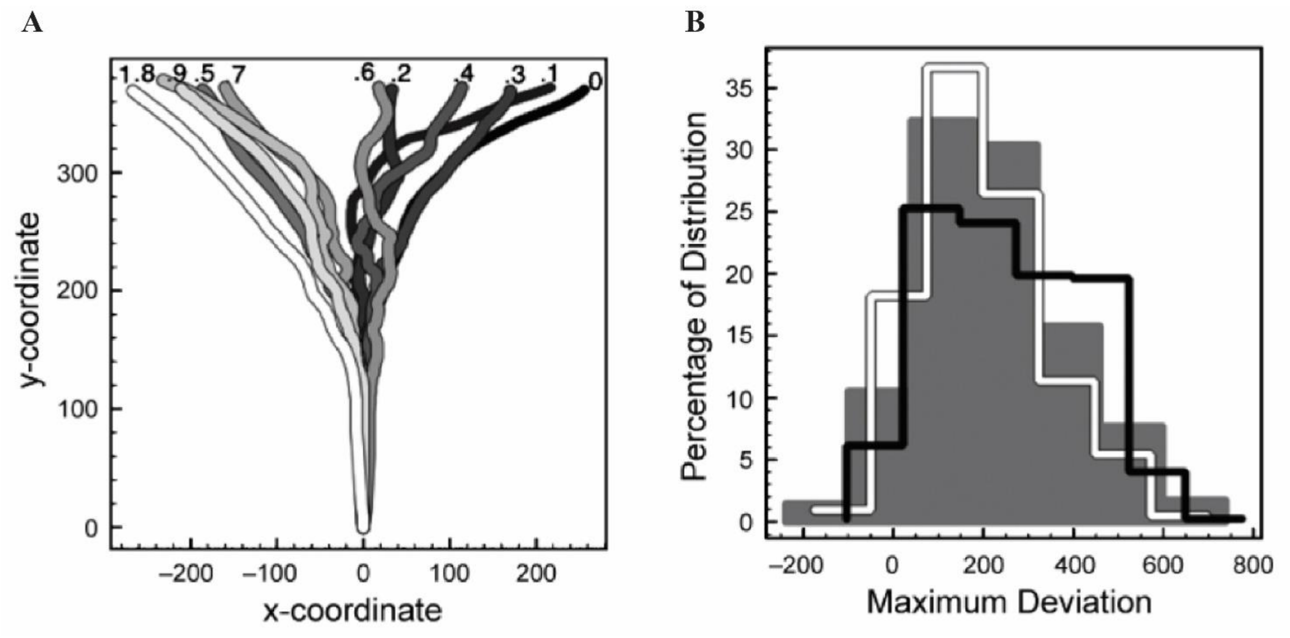

Figure 1.2 The influence of truth value on mouse-tracked trajectory curvature as indicated by **A)** The mean mouse cursor trajectories for questions with truth values ranging from high (1.0 true, white) to low (0.0 true, black) and **B)** A histogram of trajectory curvature distribution for questions with high truth value (white), low truth value (black) and middle truth value (gray). Adapted from McKinstry *et al.,* 2008.

Although reach-tracking has proven very fruitful for quantifying decision competition reflected in real-world, 3-D reaches (Chapman *et al.*, 2010a; Chapman *et al.*, 2010b; Gallivan & Chapman, 2014; Gallivan *et al.,* 2018), of particular relevance to the current study is the application of such paradigms within 2-D, computerized spaces. Computer-mouse tracking during computerized choice selection has been shown to be a highly sensitive, flexible, and scalable technique for the examination of decision processes (Freeman, 2018; Stillman, Shen & Ferguson, 2018; Hehman, Stolier & Freeman, 2015; Stillman *et al.,* 2020).

1.4 – Mouse-tracked markers of decision difficulty

Designed to provide a millisecond-level window into decision processes as they unfold over time and reveal the information-rich interplay between movement and underlying cognition, mouse tracking most notably offers opportunity to measure motoric indicators of decision difficulty. More precisely, classic mouse tracking designs are thought to compel the expression of relative accumulated evidence during decision processing in mouse movement responses (Schoemann *et al*., 2019; Stillman *et al.,* 2020; Figure 1.1). Requiring participants to start with their mouse cursor centered at the bottom of the computer screen and necessitating the selection of one of two (as is most common in reach-decision tasks) choice options located in the top left and top right corners of the screen, classic mouse tracking paradigms subject mouse movements to two primary forces: a default vertical force upwards and a horizontal force leftward or rightward proportional to the rate of evidence accumulation (Stillman *et al*., 2020). Combined, these two forces create a continuum of relatively direct or indirect trajectories, reflecting the strength of competition between choice options and thus the relative difficulty of the decision at hand. Deviations from a direct path are interpreted as demonstrating greater competitive pull towards the unchosen response, and therefore trials demonstrating greater degrees of trajectory curvature are considered to reflect more difficult decisions (see Figure 1.1 B). Given this conceptualization of choice difficulty, it follows that metrics of trajectory curvature reflecting divergence from a direct trajectory can serve as a measure of decision difficulty. The current study will use maximum absolute deviation (MAD) for this purpose (see Section 2.4.1.3 for implementation details).

1.5 – Measures of decision difficulty over time

While the power of a mouse-tracked approach lies in its ability to access the dynamic reflections of cognitive states through motoric measures, it must be noted that these measures, while sensitive to decision difficulty and the real-time evolution of decision processes as they unfold during movement time, may not provide a complete account of the cognitive competition driving a choice. Prior to movement onset, reaction time (that is, a measure of time beginning after choice presentation but terminating with motor response initiation) has also been shown to reflect cognitive conflict, with longer reaction times demonstrated by more difficult decisions (McCarthy & Donchin, 1981; Palmer, Huk & Shadlen, 2005; Rangel & Hare, 2010). A transition away from purely reaction time-based models of decision competition rose with the recognition that factors other than decision difficulty contribute to reaction time (e.g., “nondecision time” comprised of factors such as stimulus encoding and motor latency; Ratcliff & Tuerlinckx, 2002; Ratcliff & McKoon, 2008), making it difficult to extract purely difficulty-driven components.

Despite a newfound focus on motoric measures of decision difficulty, however, recent work has shown mouse-tracked trajectory-derived measures of decision difficulty to be nonredundant with reaction time (Stillman *et al.,* 2020), presenting a case in which each measure may provide unique indexes of decision processes.

The nuanced complexity of each index of decision difficulty and their relationship can be observed in experiments examining risk-based decisions pertaining to reward and loss. In a study presenting participants with a variety of gambles (e.g., a choice between a 50/50 gain/loss gamble or a certain option always equal to $0) to examine decisions involving risk, Stillman, Krajbich and Ferguson (2020) found motoric measures to be only moderately correlated with reaction time ($r = 0.32$), with the movement-derived measures being more predictive of risk preferences. Important to note, however, the authors also found that mouse-tracked measures outperform reaction time to a greater extent when predicting loss-aversion driven risk preferences (a phenomena in which loss avoidance drives preference to a greater extent than equivalent opportunity for gain) compared to risk preferences driven by diminished utility (a phenomena in which greater loss aversion is experienced for decisions involving greater opportunities for gain, even in the presence of only positive outcomes; e.g., $10 is valued less than twice as much as $5). This intricate and flexible connection between decision processes occurring prior to movement and during movement is further exemplified in an earlier study conducted by Chapman and colleagues (2015). Varying the amount of time participants spent observing two choice options before moving to enact a selection (stimulus-response intervals varying between 50 milliseconds before option presentation to 750 milliseconds after), Chapman and colleagues found that trials with less target processing time gave rise to reaches with more trajectory curvature, even when decision difficulty remained constant (i.e., the values of choice option did not change).

Together, studies such as these suggest that the manifestation of decision difficulty in reaction time and post-reaction time measures (movement trajectories and movement time) could be flexible. That is, decision processes (e.g., evidence accumulation) may be sequestered in reaction time prior to movement onset or seep largely into movement time (to be expressed in movement), and the timeframe of decision resolution can be adapted based on task demands.

The seeping of choice competition into movement time appears to be dependent on the amount of evidence accumulated prior to movement commencement, determined by factors such as time allowed for option processing prior to movement (Ghez, Gordon, Ghilardi *et al.,* 1990; Chapman *et al.,* 2015; although effects are still apparent when movements are self-initiated without time restriction; Wispinski *et al.,* 2017) as well relative strength of evidence for selection offered by a choice option (e.g., overall option value). Interestingly, the unconscious time course assignment of decision processes may also depend on the processing requirements of the choice in question (e.g., loss aversion vs. reward gain; Chapman *et al*., 2015; Stillman *et al*., 2020). Ultimately, it is perhaps only through capture of both time and motoric measures of decision making that we can obtain a comprehensive account of decision difficulty. Further, as different decision domains likely demand different comparative processes and therefore may present distinct distributions between pre- and post-movement measures, it is perhaps only through an examination of multiple indexes of decision difficulty across multiple decision tasks that we may begin to understand the relationship between these measures and their ability to accurately reflect decision difficulty.

1.6 – The current study

The primary goal of the current study was to examine how measures of decision difficulty (both pre- and post-movement) vary across decision-task, the input device recording the decision information and the platform used for study implementation. To this end, we deployed a classic mouse-tracking, reach-decision paradigm where participants were asked to complete a Numeric-Size Congruency task requiring objective perceptual judgements of which of two digits with different physical sizes was numerically larger (Faulkenberry, Cruise, Lavro & Shaki, 2016), a Sentence Verification task requiring semi-subjective conceptual judgements about the truth value of statements varying in truth value and negation (Maldonado, Dunbar & Chemla, 2019), and a Photo Preference task requiring a subjective judgement of preference between two images varying in pleasantness (Koop & Johnson, 2013). Section 1.6.2 provides additional details about these tasks. Within each task, reaction time, movement time and trajectory curvature were examined as indexes of decision difficulty (see Section 2.4.2).

1.6.1– Motivations

Discussed in relation to evident gaps in current reach-decision literature, there were three primary questions motivating the current study:

1) *How do measures of decision difficulty change across testing device: computers, tablets and smartphones?* Trajectory tracking methods have been shown to produce reliable results using 3D reach-tracking (Chapman *et al.*, 2010a; Chapman *et al.*, 2010b; Gallivan & Chapman, 2014; Gallivan *et al.,* 2018), stylus- and mouse-tracking on a large screen (Moher & Song, 2019) and 15 inch touch screen (Santens, Goossens, & Verguts, 2011), demonstrating successful application of these recording to gross motor movements (e.g. arm reaches) as well as fine motor movements (e.g. mouse movements). Despite recognition of trajectory tracking as an important and increasingly popular tool for the understanding of decisions and decision processes, however, these approaches remain relatively inaccessible and impractical outside of laboratory exploration (for example, by requiring an OptiTrack motion capture system, Chapman & Gallivan, 2014; MouseTracker or other test deployment software, Freeman & Ambady, 2010; or large computer screens). Given the rise of mobile devices as the primary way many people

access the internet, the current study aims to leverage the widespread use and familiarity of web/tablet/smartphone-based apps to test the feasibility and reliability of collecting motoric measures through these more accessible platforms. Specifically, the current study used online assessment platforms to deploy a classic trajectory tracking paradigm, comprised of the three reach-decision tasks previously introduced, across three device types of different size and user-interaction requirements: personal computers (mouse-based interactions), tablets (finger or stylus-based interactions) and smartphones (finger-, thumb- or stylus- based interactions). Importantly, driven to explore the sensitivity of capturing reach-tracked metrics of decision-making difficulty in more noisy, real-world settings, the specifications of devices used by participants were uncontrolled aside from any requirements mandated by the testing platform in use (e.g., mobile devices limited to Android tablets or smartphones using Google Chrome Web applications, see 2.2.1 for discussion).

2) *How do measures of decision difficulty relate to each other and does this change across task and device?* Measures of trajectory curvature (interpreted as a competitive pull from the non-chosen choice; Spivey, Grosjean, & Knoblich, 2005), reaction time, and movement time obtained during mouse-tracked, reach-decision tasks have been shown to provide a metric of the relative difficulty of decisions (McKinstry, Dale, & Spivey, 2008; see Section 1.4 for discussion). While these measures of decision difficulty have been demonstrated across a variety of decision domains, they are reported in different studies with different groups of participants. To our knowledge, no study has examined how decision difficulty is expressed across measures and whether the expression of decision difficulty across measures is consistent across tasks. To that end, the current study aims

to i) replicate task-specific metrics of competition between choice options in three independent decision domains differing in choice stimuli, objectivity and processing requirements using reaction time, movement time and trajectory curvature as measures of decision difficulty and ii) uncover the relationship between measures of decision difficulty and assess whether these relationships remain consistent across the decision domains and across different input devices.

3) *How does the implementation platform affect measures of decision difficulty?* Recent works examining common methodological practices (i.e., design choices) applied during mouse-tracked reach-decision tasks have shown that, among other factors, starting procedure (e.g., static vs dynamic; Schoemann *et al.*, 2019), response requirements (e.g., cursor click vs. hover; Kieslich *et al.*, 2019) and stimulus placement (near or in responses boxes located in the upper corners of the screen or distanced from response boxes; Kieslich *et al.,* 2019) influence mouse trajectory curvature and thus the strength of relationships observed between experimental manipulations and implied results (Schoemann *et al.,* 2020). Not yet examined, however, are outcome differences that may arise from platform-dependent experimental procedures. Oftentimes online assessment platforms allow for identical task designs but vary in terms of their data export profiles, requiring individualized data processing and cleaning strategies. As a component of a Mitacs Accelerate International internship in which the objective of the partner organization, Neurosight Ltd., was to validate a novel online assessment platform, the current study also aimed replicate task-specific results captured using an established online testing platform (in this case, Labvanced; see Section 2.0) using a novel platform developed by the partner organization (Horizon; commercial release pending).

Together, these three separate yet inter-reliant questions informed the objective of this study: to assess whether within-participant metrics of decision difficulty remain consistent across decision domain, data collection device, and finally implementation platform.

1.6.2 – Reach-decision tasks for replication

Informed by the objectives discussed in previous sections (see Section 1.6.1), the current study aimed to replicate three reach-decision tasks shown to sensitively reflect decision-difficulty dependent cognitive dynamics through cursor-tracked measures. Described below, each task employs a classic mouse-tracking reach-trajectory paradigm, presenting two choice options and requiring mouse-movement for option selection. Together, these tasks span a range of decision domains from objective perceptual judgments (e.g., digit discrimination), to semi-subjective conceptual judgements (e.g., truth value of a statement), and finally subjective judgements of preference (e.g., preference for one photograph over another). The variability in objectivity (e.g., objective, semi-subjective, subjective), stimulus characteristics (e.g., numeric, alphabetic, image) and processing requirements (e.g., perceptual discrimination, conceptual discrimination) offered by this selection of tasks allowed for inferences to be made about the consistency of different metrics of decision difficulty (e.g. trajectory characteristics, movement time, reaction time; see Sections 2.4.2) to be generalizable across remarkably distinct decision domain categories whose boundaries are rarely crossed in decision making literature.

1.6.2.1– Numeric-Size Congruency

The size congruency effect is characterized by an interactive processing of physical and numerical size, such that the physical size of a number can increase or decrease the ease with which it's numerical size is recognized (Henrik & Tzelgov, 1982; Faulkenberry *et al.*, 2016; Sobel & Puri, 2016). In numeric comparison tasks, participants typically demonstrate faster selection of numerically larger numbers in trials where the digits being compared are congruent in size and value (e.g., the number larger in value is also physically larger than its pair; Henrik & Tzelgov, 1982). Of interest to the current study, this effect also gives rise to differences in reach curvature between congruent and incongruent trials tested via mouse-tracking (Figure 1.3; Faulkenberry *et al.,* 2016).

In a study requiring participants to indicate with their mouse cursor which of two numbers presented at the top-left and -right of a computer screen were larger in numerical value (see Figure 1.3 A for trial design), Faulkenberry and colleagues (2016) tested 6 pairs of numerals (1–

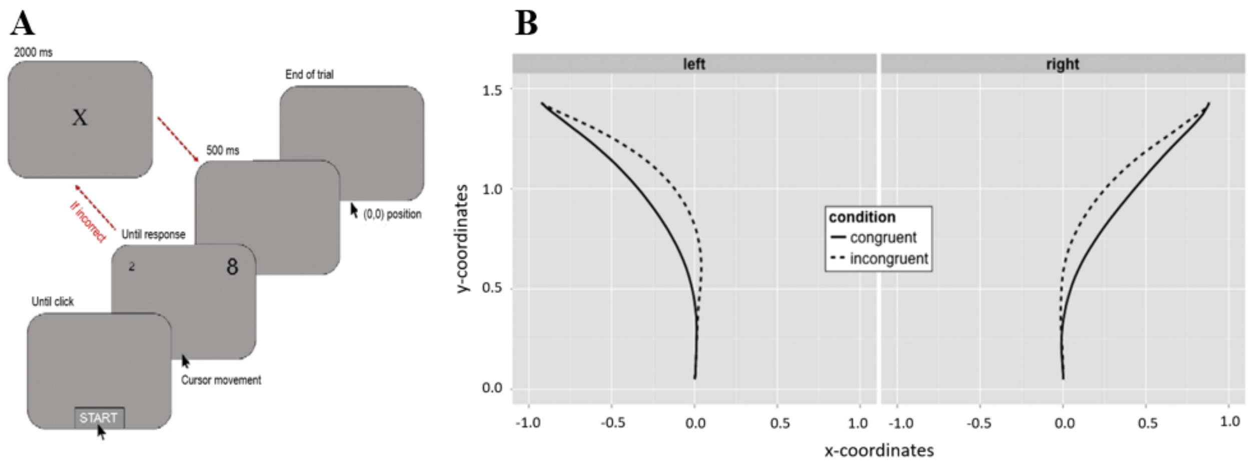

Figure 1.3 Adapted from Faulkenberry *et al.,* 2016. **A)** Trial design for a typical size-congruency task. Participants click Start to initiate trial, then made their response by moving the cursor to click on the number option of their choosing. Feedback was only provided if the response was incorrect. Shown is a congruent trial where the physically smaller digit is also smaller in numerical value. **B)** Mean left and right response trajectories for congruent and incongruent conditions, in screen coordinates.

2, 1–8, 1–9, 2–8, 2–9, and 8–9) that were physically and numerically congruent (the number larger in value was also physically larger) or incongruent (the number larger in value was physically smaller).

In addition to the expected congruency-dependent differences in temporal measures, the authors found that response trajectories for incongruent trials were significantly deflected toward the incorrect response alternative, indicating greater competitive pull relative to congruent trials (see Figure 1.3 B). Subsequent experiments indicated that this size congruency effect continues in the absence of forced movement initiation times (e.g., no specified time threshold), but is modulated by numerical distance, with the difference in curvature between congruent and incongruent trials increasing the further apart the number options are in value (Faulkenberry *et al.,* 2016).

1.6.2.2 – Sentence Verification

Tasks involving judgements about the truth value of sentences have demonstrated a bias towards affirming truth (with truth valuations proceeding more quickly than false valuations; McKinstry *et al.*, 2008) and slower reading responses when negation is present (i.e., for sentences with negative polarity; Wason, 1959). Further, an interaction between truth value and negation has been demonstrated, with negation introducing longer sentence processing times for true statements compared to false statement (Wason, 1959; Dale & Duran, 2011; Maldonado *et al.,* 2019). Seminal work by Dale and Duran (2011), demonstrated that truth value and negation, as well as their interaction, also provide predictable changes in reach-trajectories in addition to producing response timing effects. While having participants judge the truth value of simple statements presented one word at a time, the researchers found that true sentences exhibited a greater increase in trajectory curvature when negated than false sentences. Unlike prior response time studies, however, this mouse-tracking task did not elicit a main effect of truth value despite the significant main effect of negation and the truth value-negation interaction.

Illustrated Figure 1.4, Maldonado and colleagues (2019) replicated this examination of truth value and negation using a classic reach-trajectory paradigm presenting participants with complete sentences upon trial commencement (rather than paced word presentation like Dale and Duran, 2011). After presenting a simple declarative sentence, this replication study had participants reach with their mouse cursor from a bottom-centered start button to a choice option of “true” or “false” presented to the top-right and -left of the computer screen (see Figure 1.4 A). The statements presented could be true or false, negated or non-negated. In their adapted task, Maldonado and colleagues also found that affirming true negated sentences produced trajectories with more curvature compared to true non-negated sentences (see Figure 1.4 B), ultimately reproducing findings from Dale & Duran’s seminal work using a testing procedure relevant to the current study.

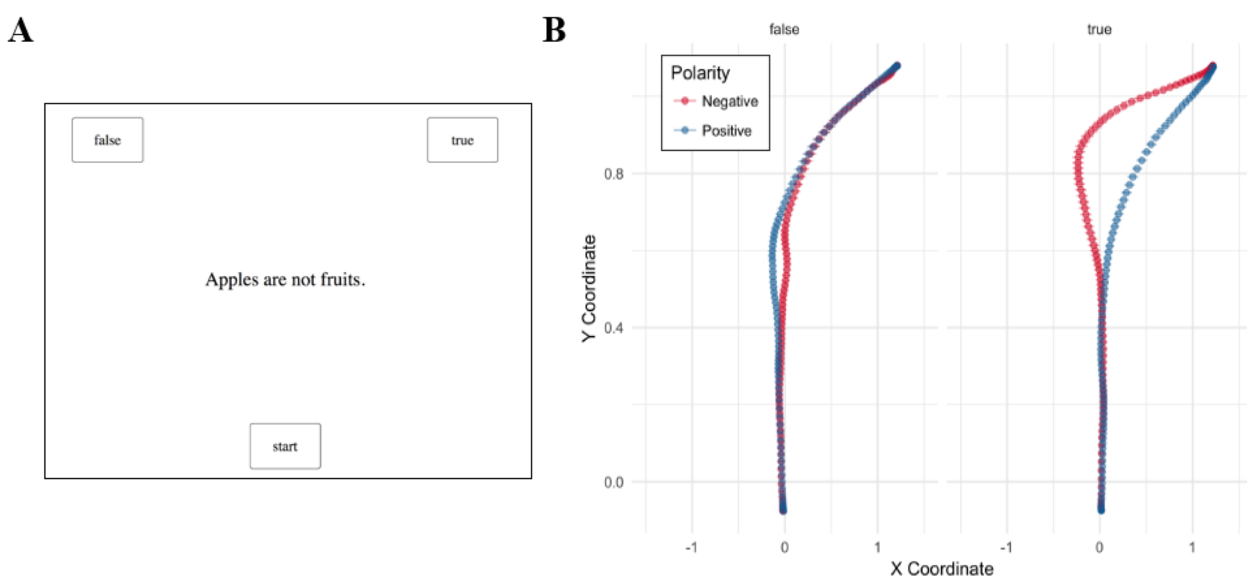

Figure 1.4 Adapted from Maldonado *et al.,* 2019. **A)** Trial design for the Dale and Duran sentence verification task replication. Participants indicated the truth value of a statement presented in the center of the screen by moving from the “start” position as the bottom of the screen to one of the two options at the top-left or top-right of the screen. Shown is a trial in which the statement is false and negated. **B)** Mean trajectories for true and false responses, in screen coordinates.

4.1.1.1 – Photo Preference

Most tasks implemented for reach-trajectory investigations of decision making involve some degree of objectivity in the determination of correct responses (e.g., McKinstry *et al.*, 2008; Dale & Duran, 2011; Faulkenberry *et al.,* 2016), with few investigating decisions in which the competitive pull towards choice options is dictated purely by subjective preference (such as snack choice; see Wispinski *et al.,* 2020 for description of unpublished data). Koop and Johnson

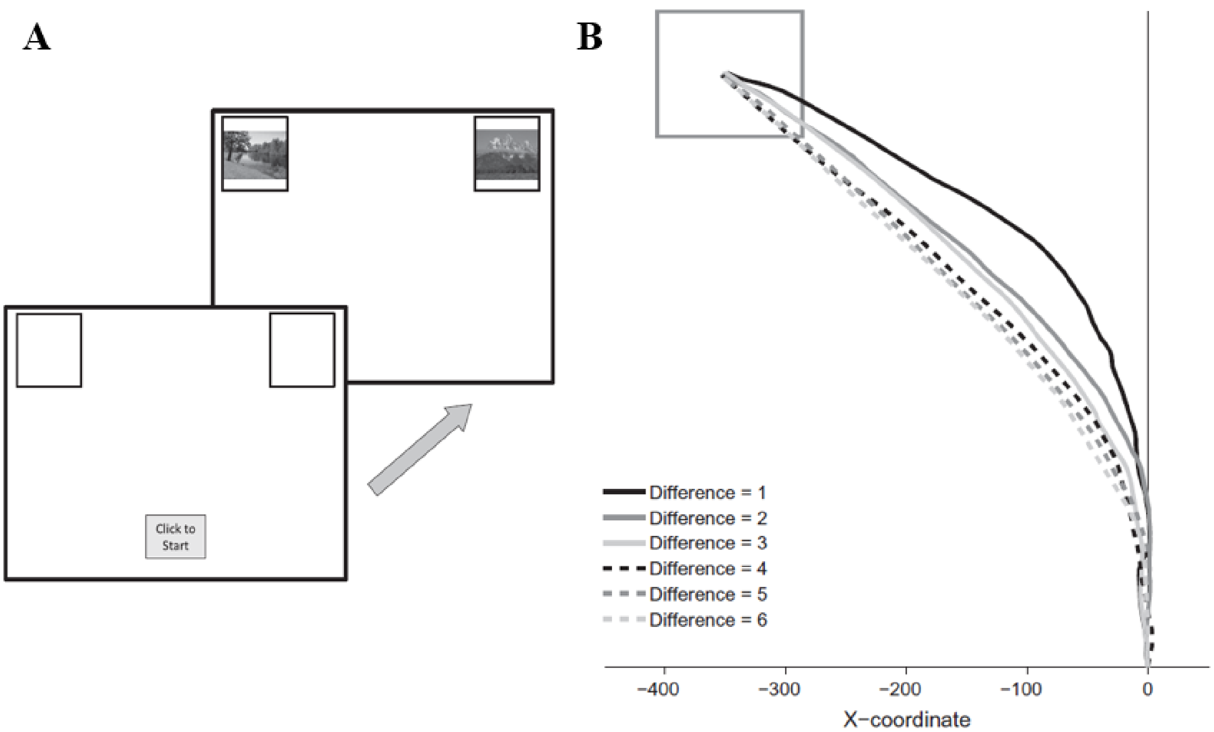

Figure 1.5 Adapted from Koop & Johnson, 2013**. A)** Photo preference trial design. Participants were initially presented with a ‘‘Start’’ button and two empty response boxes, which were populated with response options once the ‘‘Start’’ button had been clicked**.** Selection of the preferred photo could then occur. **B)** Mean response trajectories for each degree of photo pair difference, in screen coordinates.

(2013), however, set the precedent for use of traditional reach decision paradigms in examinations of subjective preference decision domains using a photo preference task. Their study simply required participants to select which of two photos they preferred (see Figure 1.5 A for task design). Photos (derived from the International Affective Picture System; Lang, Bradley

& Cuthbert, 2008) were paired according to pleasantness rating, with photo pairs ranging from similar (difference of 0) to dissimilar (difference of 6). Depicted in Figure 1.5, the reach-trajectories elicited by this task indicated a preference for the more pleasant photo choice option, and a greater competitive pull towards the unchosen alternative photo as pairs increased in pleasantness rating similarity. This established a photo preference effect in which photo pairs with large differences in pleasantness (e.g., difference of 6) produce more direct trajectories towards the more pleasant option, while photo pairs with smaller differences in photo pleasantness (e.g., difference of 1) produce trajectories with more curvature (see Figure 1.5 B).

2.0 - Materials and Methods

2.1 – Introduction

The objective of this study was to assess whether within-participant metrics of decision difficulty remain consistent across decision domain, data collection device, and finally implementation platform. To accomplish this objective, this study varied along three dimensions: task, device and platform (see Figure 2.1).

First, to assess the consistency of metrics of decision difficulty at the level of task, this study aimed to replicate three independent reach-decision paradigms previously shown to elicit decision difficulty-dependent behaviours, each of which offers a distinct decision domain

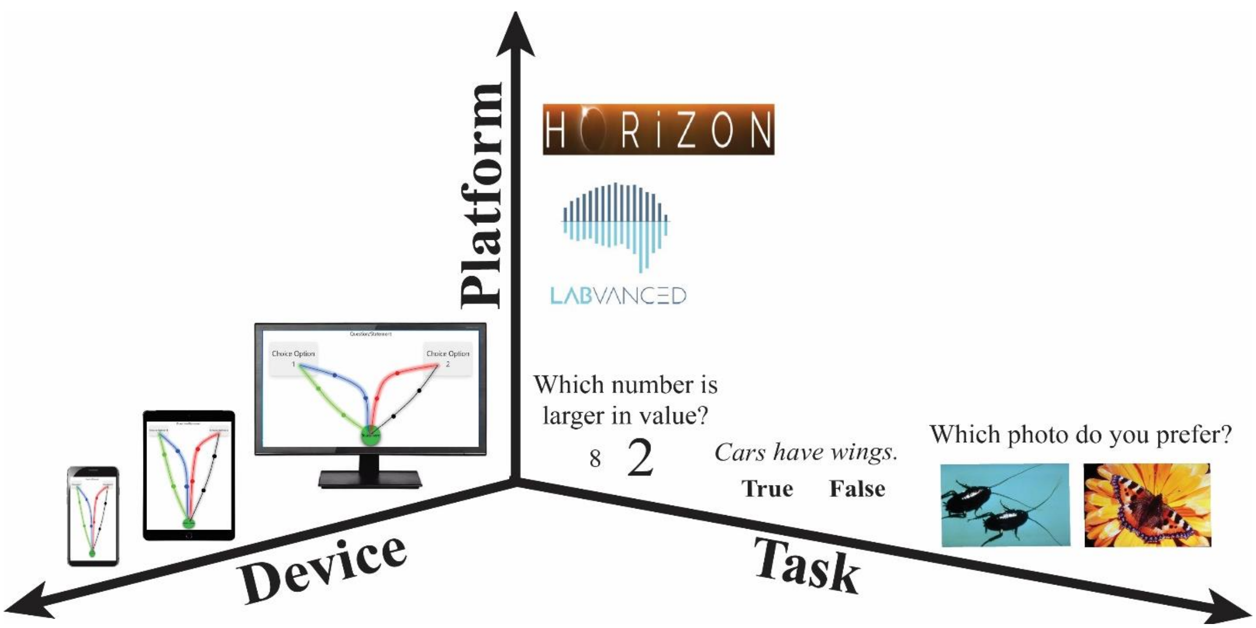

Figure 2.1 A visual representation of the three dimensions along with the current study varied.

differing in choice stimuli, objectivity and processing requirements. These tasks include a Numeric-Size Congruency task requiring objective perceptual judgements of which of two digits with different physical sizes was numerically larger (Faulkenberry *et al.*, 2016), a Sentence Verification task requiring semi-subjective conceptual judgements about the truth value of statements varying in truth value and negation (Maldonado *et al.*, 2019), and a Photo Preference task requiring subjective judgements of preference between two images varying in pleasantness (Koop & Johnson, 2013). See Section 1.6.2 for details.

Comprised of a combination of these three tasks, the study was deployed across three testing devices using Labvanced, an online, device-scalable experiment creation platform (www.labvanced.com). Specifically, the study was completed using one of three different devices: a personal computer, a tablet or a smartphone. As tests were completed using a consistent testing platform independent of the device used, this allowed comparisons to be made along the dimension of device (i.e., Device comparison, see Sections 2.5.2 and 3.1), permitting the assessment of decision-difficulty metric consistency when the testing device changes in size and user-interaction profile (e.g., mouse vs. touchscreen).

As a component of a Mitacs Accelerate International internship and motivated by the interests of the partnering organization (Neurosight Ltd.), the consistency of decision difficulty metrics across testing platform was also examined. To achieve this goal, Horizon, a second online assessment platform offering a readymade reach-decision paradigm architecture with customizable stimuli components (www.neurosight.io; commercial release pending), was introduced for study implementation using a personal computer. When compared against results obtained through the Labvanced platform also completed by personal computer users (i.e., device is held constant), this allowed for comparisons to be made along the dimension of platform (i.e., Platform comparison, see Sections 2.6.2 and 3.2).

Study design and procedure were kept consistent between platforms, with slight interface design allowances made to accommodate use of different devices within Labvanced (see Section 2.3.2). Importantly, despite identical task designs, the two platforms required unique participant recruitment (see Section 2.2), data export profiles, and data processing and cleaning strategies (see Section 2.4).

2.1.1 – Power analysis

An a-priori power analysis using computational statistics simulation methods was conducted to determine sample size based on effects estimated from sample data. For complex study designs, computer simulations are a useful alternative for estimating power (and conducting subsequent sample size calculations) when use of power equations are no longer efficient (Arnold, Hogan, Colford & Hubbard, 2011; Feiveson, 2002).

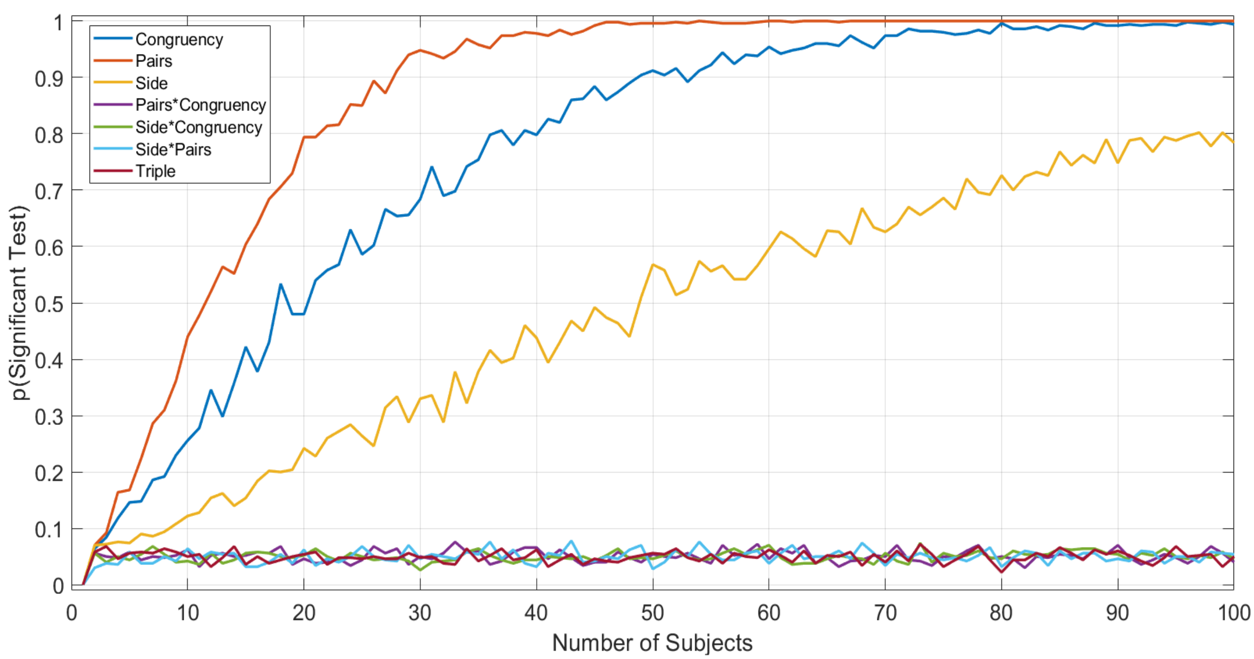

Figure 2.2 The projected power outcomes as a function of sample size for each estimated Numeric-Size Congruency effect and interaction.

A 7-participant pilot study was conducted to provide an estimate of true condition means and within- and between-subject variance. Due to time constraints, only the Numerical-Size Congruency task (as described in Section 2.3.3.1) was tested, deployed on a personal computer during in-person testing and implemented using the Horizon testing platform. As the physical-numerical size congruency and numeric value of the choice pairs have been demonstrated to drive decision difficulty effects (Faulkenberry *et al.*, 2016), these factors were the focus of our power analysis and subsequent sample size justification. Side of number presentation, on the other hand, has no prior indication of having significant effects on the curvature of reach trajectories, and is therefore disregarded in our choice of sample size (this, however, may be an inaccurate assumption. See discussion of right-hand bias in Section 4.2).

Condition mean and variance parameters generated by the pilot analysis were used to simulate novel data (assuming normally distributed noise) over many experiments with varying sample sizes. Ultimately, 49,000 experiments were simulated with sample sizes ranging from 2 to 100 participants (500 simulations per sample size). Each simulated experiment was then analyzed using a 3-factor within-subjects RM-ANOVA and the results aggregated to provide an estimate of power. Figure 2.2 illustrates the projected power outcomes as a function of sample size for each of our effect estimates. Overall, these simulations estimated a sample size of 40 participants to obtain 80% power to detect significant Numerical-Size Congruency effects at the 5% level of significance. The same number of participants will achieve 95% power for detection of significant differences between Number Pair conditions. While this proposed sample size would allow us to achieve sufficient power for our effects of interest during in-person and computer-based testing, we anticipated greater within- and between-subjects variability for online testing in uncontrolled settings and using novel devices. Further, online testing allows access to much larger sample sizes while maintaining feasibility of our projected timeline and study resources. As such, we ultimately chose to maximize our sample beyond the estimated sample size, with 40 participants recognized as a minimum required for each testing device and platform.

2.2 – Participants

In all cases, participants had no prior knowledge about the objectives or design of the experiment, and could only complete the experiment once (i.e., there were no repeat participants between devices). Participants self-reported age, gender, handedness and visual acuity, in addition to completing a brief survey about their English language proficiency, habitual activities requiring hand-eye coordination, and typical use of their chosen device for participation (see Figures 2.3 and 2.4 for a complete demographic summary). No participants were excluded based on demographic criteria.

Participants could, however, be excluded from analysis based on performance. The basis of these exclusions was insufficient (< 50%) good trials within at least one of the experimental tasks or in any of the unique task conditions (whether due to unreliable data capture, trials unrepresentative of typical behaviour, or incorrect responses; see Section 2.4.3 for further details on data cleaning procedures). Device- and platforms-specific participation is discussed below.

All experimental procedures were approved by the University of Alberta's Research Ethics Office.

2.2.1 – Labvanced

Three-hundred and five participants were recruited online through Amazon's Mechanical Turk (mTurk; http://www.mturk.com), a participant recruitment system shown to produce reliable respondents in other studies (including those employing reach-decision paradigms, e.g., Dale and Duran, 2011), for participation using the Labvanced platform. Three different groups of Labvanced participants completed the experiment using a personal computer, a tablet, or a smartphone of their choosing. The devices used were uncontrolled except for requiring use of a separate mouse (wired or wireless) during computer use, or an Android operating system and touch-screen device interaction (via finger, thumb or stylus) during tablet or smartphone use (note Apple iPads and iPhones were not compatible with cursor tracking in the Labvanced platform).

2.2.1.1 – Personal Computer

A total of one hundred and one participants completed the study using a personal computer. Of those, nine were excluded after participation for not meeting device interaction requirements (i.e., use of a wired or wireless mouse). Data pre-processing and cleaning (Sections 2.4.1.1 and 2.4.3, respectively) resulted in the exclusion of a further nine computer users, resulting in data from 83 computer users (25 female, 56 male, and 2 who preferred not to say; $M_{age} = 33.75$, $SD_{age} = 9.35$; see Figure 2.3 A) being used for analysis.

2.2.1.2 – Tablet

One hundred and one participants completed the study using a tablet. Four were immediately excluded from analysis for not meeting device interaction requirements (i.e., finger-, thumb- or stylus-based interactions). Data pre-processing and cleaning (Sections 2.4.1.1 and 2.4.3, respectively) resulted in the exclusion of a further nineteen tablet users, leaving data from 78 tablet users (27 female, 50 male, and 2 nonbinary; $M_{age} = 33.51$, $SD_{age} = 6.22$; see Figure 2.3 B) to be included in the analysis.

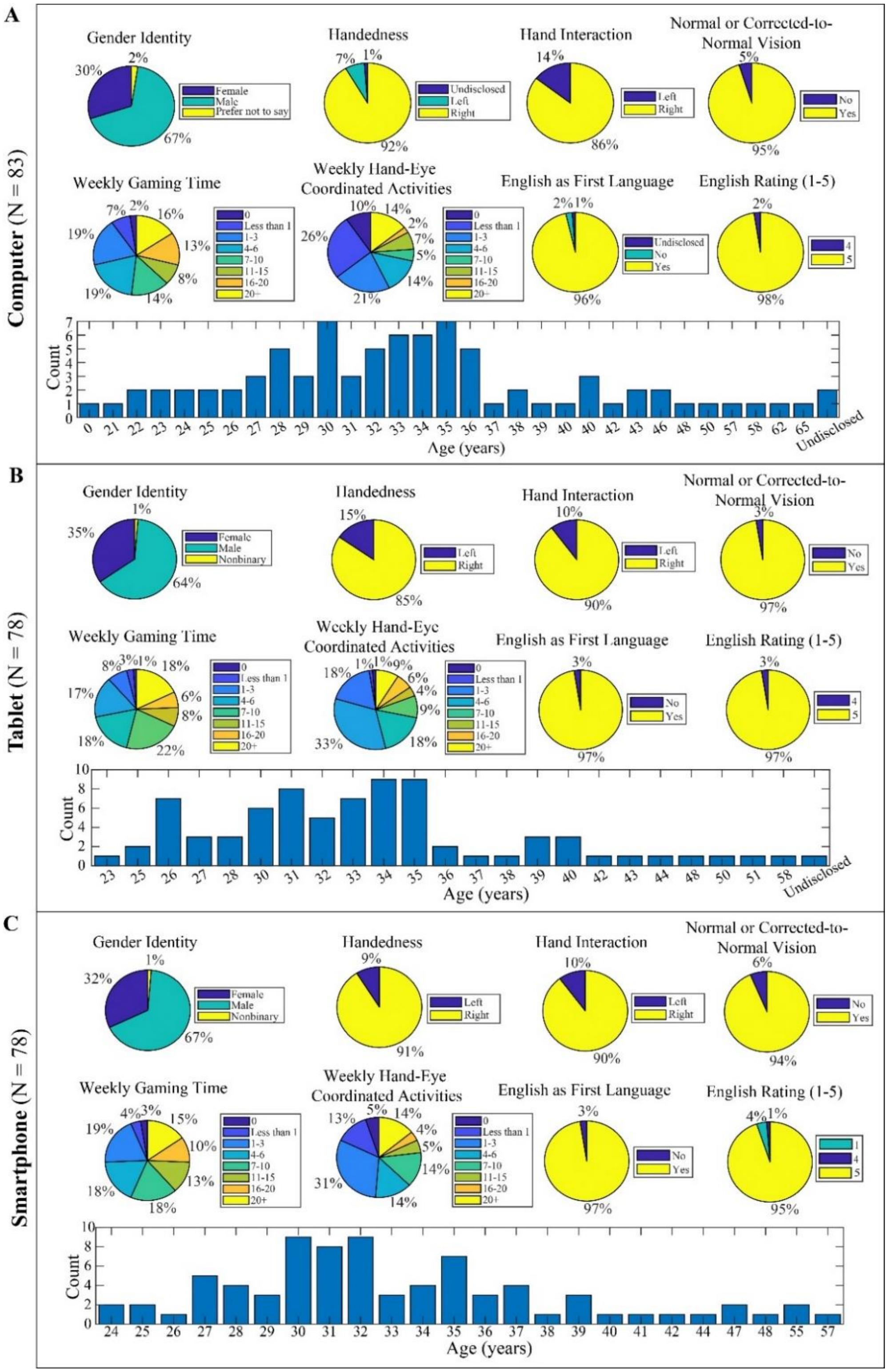

Figure 2.3 Self-reported demographic information for Labvanced participants included in the analysis, with **A)** Computer users (N = 83), **B)** Tablet users (N = 78) and **C)** Smartphone users (N = 78) shown separately. A complete summary of demographic survey questions and responses can be found in Appendix A.1.

2.2.1.3 – Smartphone

A total of one hundred and three participants completed the study using a smartphone. Of those, twenty-five were excluded based on data pre-processing and cleaning criteria (Sections 2.4.1.1 and 2.4.3, respectively). After exclusions, 78 smartphone users (25 female, 52 male, and 1 who preferred not to say; $M_{age} = 33.73$, $SD_{age} = 6.72$; see Figure 2.3 C) remained for analysis.

2.2.2 – Horizon

2.2.2.1 – Personal Computer

Fifty participants were recruited via university-wide email advert for study participation using the Horizon platform. Criteria for participation included only access to a personal computer with use of a wired or wireless mouse (laptop touchpads or trackpads were prohibited for device interaction). Four participants were excluded after participation for not meeting device

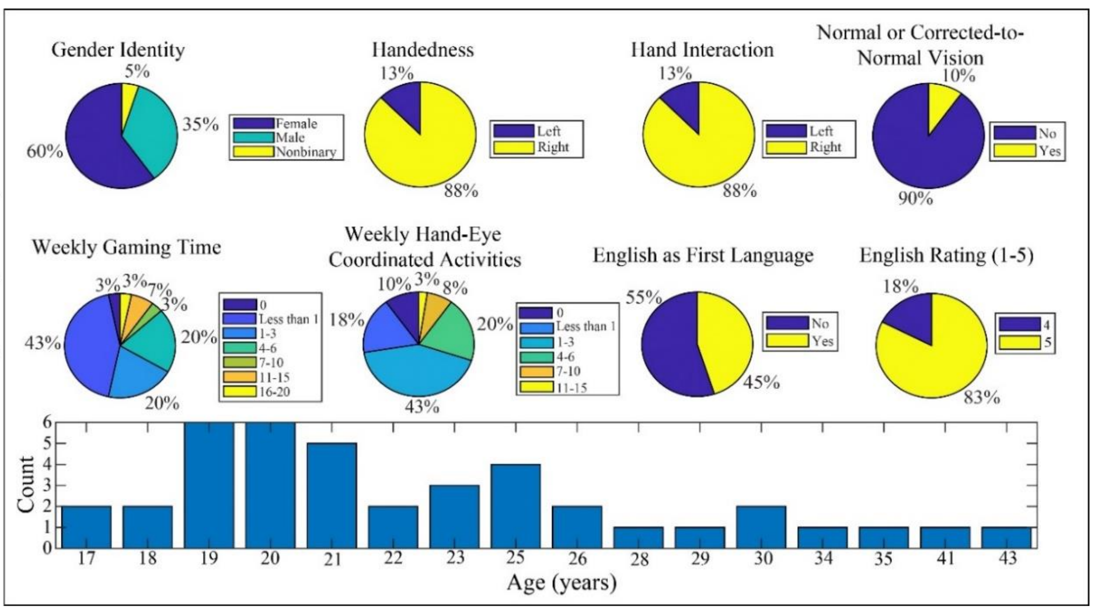

Figure 2.4 Self-reported demographic information for Horizon participants included in the analysis. A complete summary of demographic survey questions and responses can be found in Appendix A.1.

interaction requirements and data pre-processing and cleaning (Sections 2.4.1.2 and 2.4.3, respectively) resulted in the exclusion of a further six participants from the analyses. After exclusion, 40 (24 female, 14 male, 2 nonbinary; M_{age} = 23.6, SD_{age} = 6.12; see Figure 2.4) data sets remained for analysis.

2.3 – Procedure and design

2.3.1 – Procedural overview

Participants completed the experiment online using a Google Chrome browser, in a remote setting of their choosing.

Upon study enrollment, participants were provided with an electronic consent form delineating their rights, the risks of the study (of which there are none known) and the removal of any personal identifiers linked to the data collected. Electronic consent indicating voluntary participation was obtained prior study commencement. Only consenting participants gained access to the study.

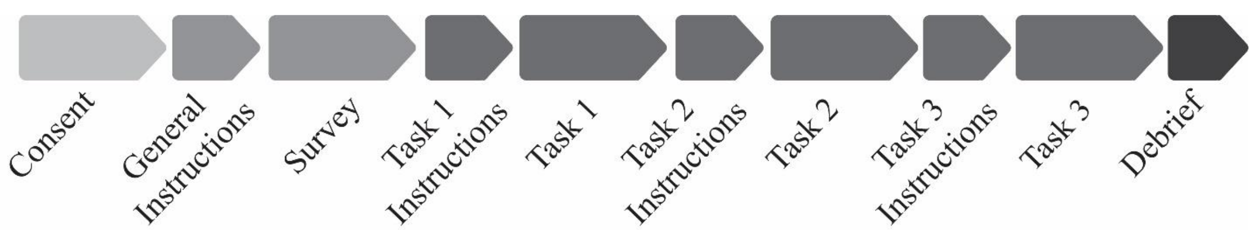

Figure 2.5 Overview of the study design, beginning with Consent (left, lightest grey) and concluding with a Debrief (right, darkest grey).

The study consisted of a survey followed by three experimental tasks (see Figure 2.5 for an organizational overview of the study). The survey asked participants to share their demographic details (e.g., age, gender, handedness, English language proficiency, visual acuity), habitual partaking in activities requiring hand-eye coordination (e.g., videogame or sport activities), the unique specifications of the device they are using for participation (e.g., brand, model, size, input device) and their typical interactions with that device (e.g., weekly use for school, work, gaming, communication). Specific survey questions and response summaries can be found in the

Appendix (A.1). The subsequent three experimental tasks each presented one of the reach-decision tasks described previously: a Numeric-Size Congruency task (adapted from Faulkenberry *et al.,* 2016; see Section 2.3.3.1), a Sentence Verification task (adapted from Dale & Duran, 2011 and Maldonado *et al.,* 2019; see Section 2.3.3.2) and a Photo Preference task (adapted from Koop & Johnson, 2013; see Section 2.3.3.3). Participants were instructed to complete the study in its entirety in a single session and were provided with detailed instructions outlining the study tasks prior to the start of each test task.

All three experimental tasks presented a classic reach-decision paradigm requiring participants to choose one of two stimuli presented at the top left and top right of their device screen. Each trial first presented a green circular start button labelled "Touch here" on the bottom center of the screen, requiring participants to navigate their mouse cursor (in the case of computer use) to or place their finger, thumb, or stylus (in the case of tablet or smartphone use) on the button to trigger the trial start. Touching of the start button triggered a three second countdown, centered on the display screen (Figure 2.6). Removal of the mouse cursor, digit or stylus from the start button or the surface of the screen would cause the countdown to pause until touch-contact within the start button had been re-established. For Numeric-Size Congruency and Photo Preference tasks, countdown onset was accompanied by a question specific to the task type appearing centered at the top of the display (Figure 2.6). Once the countdown was complete, two choice boxes appeared at the upper-left and upper-right of the screen, each presenting trial-specific choice options. Unlike the other two tasks, the Sentence Verification task instead presented the two choice options coincident with countdown onset and presented a statement (rather than a question) centered at the top of the screen upon countdown completion (Figure 2.6). Task-specific choice stimuli and trial design are described in Section 2.3.3. Participants

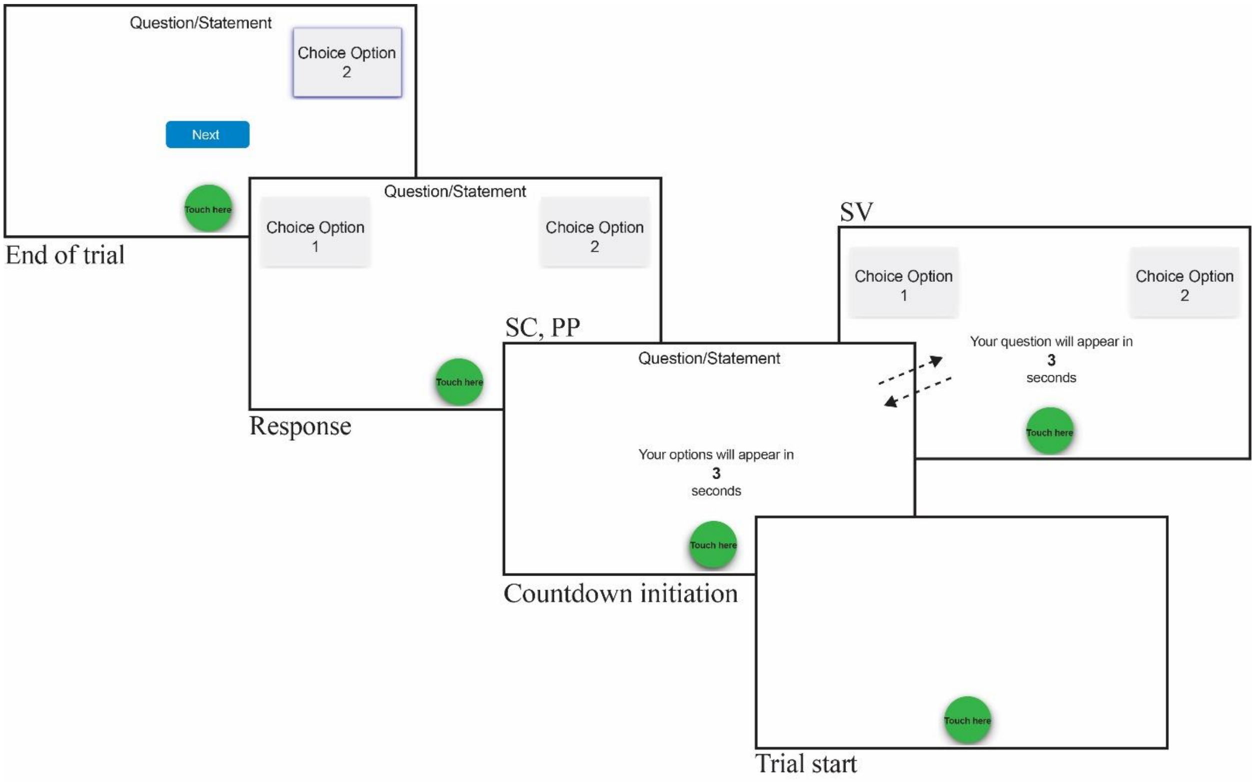

Figure 2.6 Study testing interface and experimental task trial design. The four panels from bottom-right to top-left demonstrate the trial process for Numeric-Size Congruency (SC) and Photo Preference (PP) tasks. The order of choice/question component presentation is reversed Sentence Verification (SV) tasks, indicated by the alternative second panel.

were free to select either choice option in response to the question or statement on the screen immediately upon countdown completion, with no time constraints. In the case of computer use for study participation, choice selection required participants to move their mouse cursor inside the choice box of their choosing (no mouse click was required). In the case of tablet or smartphone use, participants were required to slide their finger, thumb, or stylus across the screen to touch their selected choice box, keeping contact with the screen at all times. If touchscreen contact was lifted, an error message would appear on the screen, reading “Your finger was lifted from the screen as you moved, and we were unable to track the movement. Please touch your option now and remember in the future to keep your finger on the screen.” The choice options glowed blue to confirm contact and selection. Once a selection was made, the start button and unchosen choice option disappeared, and a "Next" button appeared centered on the screen. Participants were then free to click or press on the "Next" button to continue to the next trial, allowing them to self-pace the experiment. Figure 2.6 provides a visual representation of the testing interface and trial design.

Each experimental task presented 84 trials, taking approximately 15 minutes each to complete. Trials were randomized within each task to mitigate within-task practice effects. Similarly, the order of task presentation was counter-balanced across participants to mitigate between-tasks practice effects. Participants were encouraged to take short breaks between tasks but had a maximum time limit of ninety minutes to complete the study.

Upon completion of the three testing tasks, participants were thanked for their time and given access to a debriefing form detailing the purpose of the study. Participants recruited locally through university adverts were compensated with a $10 CAD gift-card upon study completion, while participants recruited via Amazon Mechanical Turk were compensated with $7 USD.

2.3.2 – Device specific design

The dimensions of the testing interface scaled according to the screen size of the device in use, presenting a landscape orientation for computer-based participation and a portrait orientation for touchscreen-based participation. Of note, start button, choice option and question or statement font size interface components scaled to maintain consistent component-screen proportions independent of device screen size (i.e., the proportion of interface space occupied by

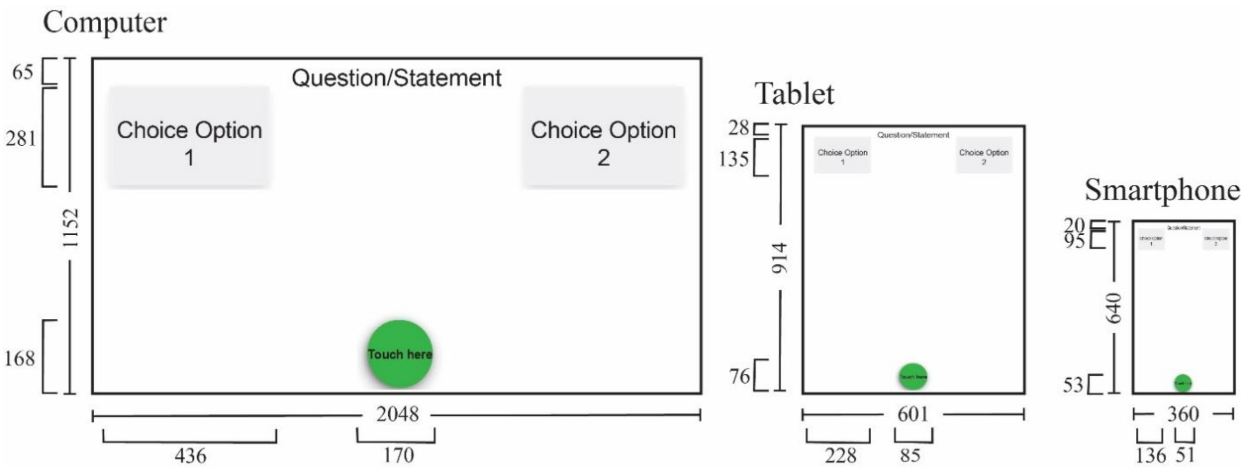

Figure 2.7 A comparison of interface arrangements between devices. Shown are representative examples of a Computer, Tablet and Smartphone testing interface. All values are reported in pixels. Specific sizes of device screens and interface components observed by participants were dependent on the size of the device used, but screen to interface component proportions remained constant within each device category. For full details regarding participant-reported screen sizes, see Table A.1.2.

each component stayed constant). Specifically, for computer-presented tests the start button occupied 1.2% of the testing interface (width = 8.3% of total screen width, height = 14.6% of total screen height), the choice options each occupied 5.2% of the testing interface (width = 21.3% of total screen width, height = 24.4% of total screen height), and the font size of the written questions or statements was set at 5.6% of the screen height. For tablet- or smartphone-presented tests, the start button also occupied 1.2% of the testing interface (width = 14.1% of total screen width, height = 8.3% of total screen height), the choice options each occupied 5.2% of the testing interface (width = 37.9% of total screen width, height = 14.8% of total screen height), and the font size of displayed questions or statements was set at 3.1% of the screen height. Examples of interface design and component sizing for each of the three devices permissible for participation use is shown in Figure 2.7.

2.3.3 – Stimuli and task-specific procedure

2.3.3.1 – Numeric-Size Congruency

The Numeric-Size Congruency task in the current study was designed to replicate Faulkenberry, Cruise, Lavro and Shaki's (2016) experiment examining the dynamics of the size congruency effect (see Section 1.6.2.1). Due to time constraints, the stimuli presented in the current study were limited to a subset of those presented in the original experiment.

	Congruent		Incongruent	
Pairs	Left	Right	Left	Right
1 - 2	2 - 1	1 - 2	2 - 1	1 - 2
2 - 8	8 - 2	2 - 8	8 - 2	2 - 8
8 - 9	9 - 8	8 - 9	9 - 8	8 - 9

Figure 2.8 A representation of the twelve Numeric-Size Congruency conditions obtained by crossing selected Number Pairs (Pairs) by Congruency (Congruent or Incongruent) and side of presentation (Left or Right). Condition design is replicated from Faulkenberry *et al.*, 2016.

For each Numeric-Size Congruency trial, the question "Which number is larger in value?" appeared coincident with the onset of the countdown timer and centered at the top of the screen (Figure 2.6). Following countdown termination two numbers were displayed simultaneously in choice boxes at the upper left and right corners of the screen, at which time participants could move to select their preferred choice. Stimuli consisted of the Arabic numerals 1, 2, 8 and 9 displayed in Arial font and presented in pairs of different physical size. From these, six choice-pair options were generated: 1 – 2, 2 – 8 and 8 – 9, with each pair either congruent in physical and numeric size (the numerically larger numeral appearing physically larger than its paired counterpart, e.g., 2 – 8), or incongruent in physical and numeric size (the numerically larger numeral appearing physically smaller than its paired counterpart, e.g., 2 – 8; see Figure 2.8). Within each condition, the numerically larger number was presented equally often on the left and the right, counterbalancing side of space effects. The twelve conditions obtained by crossing each number pair with physical-numerical size congruency and side of presentation were presented 7 times for a total of 84 trials (see Figure 2.8).

2.3.3.2 – Sentence Verification

The Sentence Verification task in the current study was designed to replicate Maldonado, Dunbar and Chemla's (2019) adaptation of Dale and Duran's (2010) linguistic negation experiment (see Section 1.6.2.3). Important to note, statement stimuli (see Appendix A.2) presented in the current study were developed from example statements provided by the authors of both previous works but exact replication of previous stimuli was not possible as no comprehensive list of statements used were published.

On each Sentence Verification trial, the choice boxes appeared coincident with the onset of the countdown timer and were populated with "True" and "False" response options in the top-left and top-right corners of the screen, respectively (Figure 2.6). Following countdown termination, a statement was displayed at the top-center of the screen, prompting participants to judge whether it was true or false by selecting one of the options that appeared earlier. Statement stimuli consisted of 21 simple declarative statements manipulated in truth value (true, false) and polarity (positive, negative). Sentence polarity was determined by the presence of negation, where non-negated sentences are considered positive in polarity (e.g., “giraffes are tall”) and negated sentences are considered negative in polarity (e.g., “giraffes are not tall”). Truth value was then manipulated by changing the adjective at the end of the sentence (e.g., “giraffes are not short” is true, while “giraffes are not tall” is false). Like with Maldonado *et al.* (2019), these two factors were crossed to generate four sentence conditions where each sentence could be a true or false statement in either negative or positive forms (exemplified in Figure 2.9; see Appendix A.2 for a complete list of generated statement stimuli). Participants saw all four instances of each model sentence in random order for a total of 84 trials.

Truth Value	Polarity	Example
True	Positive	Cars have tires.
	Negative	Cars do not have wings.
False	Positive	Cars have wings.
	Negative	Cars do not have tires.

Figure 2.9 A representation of the four Sentence Verification conditions obtained by crossing Truth Value (True or False) with Polarity (Positive or Negative). Condition design is replicated from Maldonado *et al.*, 2019.

2.3.3.3 – Photo Preference

The Photo Preference task in the current study was designed to replicate Koop and Johnson’s (2013) experiment examining the dynamics of preferential choice (see Section 1.6.2.3). Again due to time constraints, the current study limited presented photo stimuli to a subset of those presented previously. Specifically, while Koop and Johnson presented a range of stimuli ranging from unpleasant (pleasantness = 1.66) to very pleasant (pleasantness = 8.34) and simply paired stimuli based on similarity such that pairs that ranged between similar (pleasantness different = 0) and dissimilar (pleasantness difference = 6), the current study instead grouped photos into three categories of pleasantness (high, average and low; see for further detail) designed to span a similar range of pleasantness, and paired photos based on those groupings (see Figure 2.10 for stimuli overview).

For each Photo Preference trial, the question “Which photo do you prefer?” appeared centered at the top of the screen with countdown initiation (Figure 2.6). Following countdown termination two images were then simultaneously displayed in the choice boxes to the upper left and upper right corners of the screen. As in Koop and Johnson (2013), the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) was used to develop a stimulus set of paired images. The IAPS provides researchers with a standardized collection of over 1000 photographs, each normed in terms of pleasantness (affective valence), arousal, and dominance. These normative ratings were obtained in 18 separate studies, resulting in approximately 100 independent ratings per image, resulting in a nine-point scale for each affective dimension. The affective norms of interest in the current study were limited to pleasantness and arousal, given the prior validation of pleasantness as an analog to photo preference, given equal levels of arousal (Koop & Johnson, 2013).

The average ratings of pleasantness and arousal were used to select 168 pictures from the IAPS, which were then categorized as being high in pleasantness (pleasantness rating between 7 and 8), average in pleasantness (referred to as Med; pleasantness rating between 4.50 and 5.50) or low in pleasantness (pleasantness rating between 2 and 3). Images scoring higher than 8 or lower than 2 in pleasantness, and/or greater than 6.15 in arousal were excluded to minimize graphic content deemed inappropriate for the target participant pool. Selected pictures were then matched to provide all pairwise comparisons of pleasantness category, with arousal ratings held

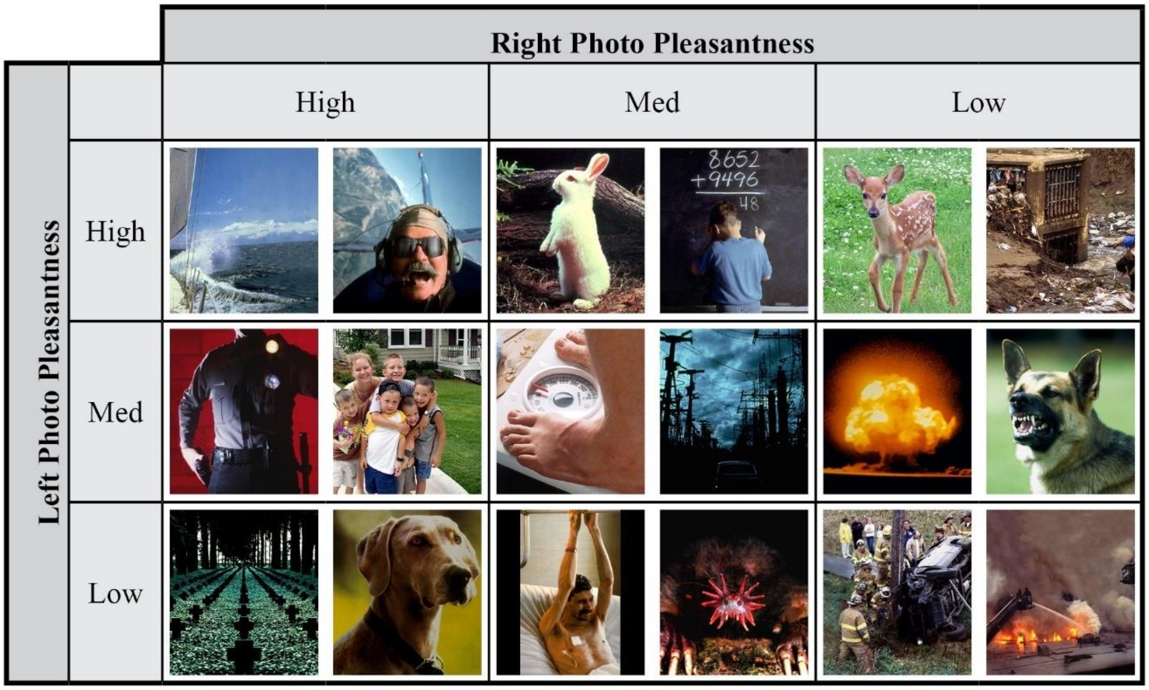

Figure 2.10 A representation of the nine Photo Preference conditions resulting from all pairwise combinations of High, Med, and Low photo pleasantness categories. As depicted above, pairs not matched in pleasantness (High – Med, High – Low, Med – Low) were counterbalanced for side of presentation. Not shown is that photo pairings matched in pleasantness (High – High, Med- Med, Low – Low) appeared twice as often to maintain equal presentations of each pleasantness pairing. Condition design is adapted from Koop & Johnson, 2013.

constant between pairs (difference < 0.30). Pairs that were not matched in pleasantness (e.g., High – Med, High – Low, Med – Low) were counterbalanced for side of presentation, while pairs matched in pleasantness (e.g., High – High, Med – Med, Low – Low) appeared equally as often as the unmatched conditions when ignoring side of space (see Figure 2.10 for examples). This allowed for 14 presentations of each pleasantness pairing (7 of each unmatched pairing for each presentation side and 14 for matched pairings), for a total of 84 trials.

2.4 – Data treatment

2.4.1 – Trajectory extraction

2.4.1.1 – Labvanced

Labvanced-acquired data treatment and measure extraction occurred using our Gaze and Movement Analysis (GaMA) platform, a custom software solution developed in-lab for streamlined gaze and movement analysis (https://www.ksr.ualberta.ca/acelab/?page_id=161).

Raw movement data acquired through Labvanced was reported at device- and server latency-dependent framerates ranging between 5 and 30 milliseconds per frame ($M_{framerate}$ = 11.66, 9.78 and 9.94 for computer-, tablet- and smartphone-acquired data, respectively). Raw movement data was therefore first resampled to 60 Hz, then filtered using a 10 Hz lowpass filter. Reach onset was then defined as the first time the mouse cursor (computer-use) or finger/thumb/stylus (touchscreen tablet- or smartphone-use) ascended to 5% of its peak velocity within the start button and after countdown had terminated. Should this velocity threshold not be achieved prior to leaving the start button, this threshold was iteratively reduced to 95% of its value until a reach onset could be defined. Reach offset was similarly defined as the first time the mouse cursor (computer-use) or finger/thumb/stylus (touchscreen tablet- or smartphone-use) velocity descended below a velocity threshold of 5% peak velocity while within one of the two choice option boxes, with this threshold iteratively increasing by 5% until reach offset could be defined within those location parameters. Reach trajectories containing fewer than seven unique data points (e.g., having 100 milliseconds or less of data) were considered to have insufficient data for analysis. Once the reach trajectory was extracted, the trajectory curvature was operationalized as described in Section 2.4.1.3.

2.4.1.2 – Horizon

Horizon-acquired data treatment and measure extraction occurred exclusively through custom processing algorithms using MATLAB (www.mathworks.com/products/matlab). As with Labvanced-acquired data, raw movement data acquired using the Horizon platform expressed sampling rates varying between 0 and 1057 milliseconds per frame ($M_{framerate}$ = 10.03), and was therefore resampled at 60 Hz prior to being filtered using a 10 Hz lowpass filter. Reach onset was defined at the as the first time the mouse cursor velocity ascended past a velocity threshold of 200 pixels per second. Reach offset was automatically defined by the Horizon platform as the first data point recorded within one of the choice option boxes. Unlike Labvanced-acquired data, Horizon-acquired reach trajectories were also space-normalized, a process involving resampling of the y-coordinate vector into a specific number of equal-space steps and computing, using linear interpolation, corresponding x coordinate and time vectors (Gallivan *et al.*, 2011; Chapman & Goodale, 2010; Chapman *et al.*, 2010a; Chapman *et al.*, 2010b). In the current case, raw trajectories were normalized to 100 equally spaced points along the vertical axis between the start position to the final position, with each point corresponding to one percent of the y-distance travelled. Reach trajectories containing fewer than seven unique data points (e.g., having 100 milliseconds or less of data) were considered to have insufficient data for analysis. Once the reach trajectory was extracted and normalized, the operationalization of trajectory data occurred in the same manner as Labvanced-acquire trajectories (see Section 2.4.1.3).

2.4.1.3 – Operationalization of trajectory data

Independent of data source or treatment method, raw trajectory data was operationalized in the same manner. Within each trial, absolute measures of deviation (Euclidean distance) of the observed trajectory relative to an ideal response trajectory (a straight line connecting the trajectory start and end position) was calculated at each x,y coordinate data point (derived from the filtered trajectories described in Section 2.4.1.1 for Labvanced-acquired data, and the 100 space-normalized bins described in Section 2.4.1.2 for Horizon-acquired data). The degree of

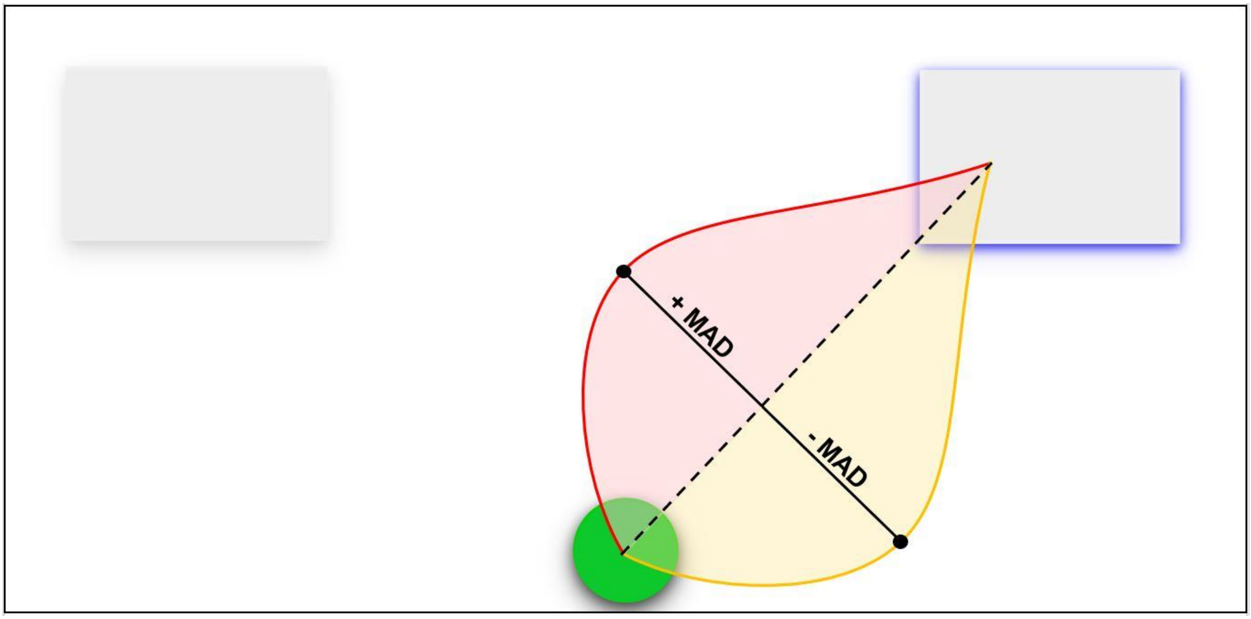

Figure 2.11 For an example response trajectory, the maximum absolute deviation (MAD) is depicted (black line) as the maximum perpendicular deviation of the trajectory from a straight line connecting the start and end points of the trajectory (dashed line). Shown are examples of two reaches opposite in MAD value (positive in red and negative in yellow) given the same start and end point (i.e., same ideal trajectory).

curvature was indexed as the maximum absolute deviation (MAD), or the measure that maximizes the perpendicular distance between the idealized and observed paths (see Figure 2.11 for an example). The computed MAD values were then interpreted as levels of deviation towards the unselected alternative option, with greater trajectory curvature towards the alternative option giving rise to larger positive MAD values, straight trajectories producing values approaching zero, and trajectories deviating away from the alternative option producing negative values. In the context of decision making, the maximum deviation of a reach was then interpreted as an estimate the level of competition or indecision between two choice targets in space, with greater positive deviations indicating greater decision difficulty (Spivey, 2008; McKinstry *et al.*, 2008; Freeman & Ambady, 2010; Koop & Johnson, 2013).

2.4.2 – Dependent measures

For each trial, the following behavioural measures were obtained:

Reaction time (seconds): time from countdown termination to reach onset.

Movement time (seconds): time from reach onset to reach offset (choice selection).

Maximum Absolute Deviation (MAD): maximum deviation of the reach trajectory from an ideal response trajectory, computed from the post-processing x,y coordinates of mouse/finger/thumb/stylus position on the device screen during choice selection (see Section 2.4.1.3).

Within-participant and within-task z-scores were computed for each dependent measure (reaction time, movement time, MAD). This standardization of within-participant measures allows for between-task and between-participant comparisons while controlling for participant variability and individual reach patterns. All analyses were conducted on these standardized values.

2.4.3 – Data cleaning

Data cleaning processes were identical independent of platform or device used for testing and were conducted using customized MATLAB cleaning scripts.

Errors on each trial could be a combination of reaches with recording errors ($\text{Computer}_{\text{Horizon}}$: M = 0.29%, Range: 0% - 16.1%; $\text{Computer}_{\text{Labvanced}}$: M = 0.73%, Range: 0% - 22.6%; Tablet: M = 4.5%, Range: 0% - 96.8%; Smartphone: M = 4.3%, Range: 0% - 92.9%), reaches with insufficient data points (see Section 2.4.1, $\text{Computer}_{\text{Horizon}}$: M = 3.7%, Range: 0% - 65.87%; $\text{Computer}_{\text{Labvanced}}$: M = 1.65%, Range: 0% - 59.1%; Tablet: M = 7.1%, Range: 0% -

90.1%; Smartphone: M = 6.2%, Range: 0% - 86.1%), reaches with "out of bounds" start or end positions ($Computer_{Horizon}$: M = 1.2%, Range: 0% - 36.1%), reaches with reaction times greater than 0.1, ($Computer_{Horizon}$: M = 0.1%, Range: 0% - 2.8%; $Computer_{Labvanced}$: M = 0.53%, Range: 0% - 6.8%; Tablet: M = 0.85%, Range: 0% - 18.3%; Smartphone: M = 0.7%, Range: 0% - 7.1%), > 3 SD of mean movement time errors ($Computer_{Horizon}$: M = 2.1%, Range: 0% - 3.6%; $Computer_{Labvanced}$: M = 1.8% , Range: 0.4% - 3.2%; Tablet: M = 1.6%, Range: 0% - 2.8%; Smartphone: M = 1.6%, Range: 0% - 3.2%), and > 3 SD of reaction time errors ($Computer_{Horizon}$: M = 1.5%, Range: 0% - 2.8%; $Computer_{Labvanced}$: M = 1.5%, Range: 0% - 4.4%; Tablet: M = 1.6%, Range: 0% - 3.2%; Smartphone: M = 1.4%, Range: 0% - 3.9%). For Numeric-Size Congruency and Sentence Verification tasks, incorrect trials were also removed from analysis (Numeric-Size Congruency: $Computer_{Horizon}$: M = 2.4%, Range: 0% - 83.4%; $Computer_{Labvanced}$: M = 2.1%, Range: 0% - 42.9%; Tablet: M = 0.83%, Range: 0% - 42.9%; Smartphone: M = 0.91%, Range: 0% - 42.9%; Sentence Verification: $Computer_{Horizon}$: M = 1.5%, Range: 0% - 2.8%; $Computer_{Labvanced}$: M = 3.3%, Range: 0% - 45%; Tablet: M = 3.6%, Range: 0% - 42.9%; Smartphone: M = 3.3%, Range: 0% - 45%). As these tasks previously demonstrated very high levels of accuracy (Dale & Duran, 2011, Faulkenberry *et al.,* 2016), incorrect responses were considered to arise from participant error, with sustained performance errors indicating participant unreliability.

In total, participants whose data was analyzed had a mean of 95.6% usable trials for analysis (Range: 83.7% - 98.4%). These trial rejection numbers fall within expected rates for reach-decision experiments (Gallivan & Chapman, 2014). Participants were rejected from further analysis if the number of errors or incorrect trials resulted in <50% correct, usable trials within any of the unique conditions within the three experimental tasks. This criterion was enforced to ensure participants had at least four trials in all conditions for analysis. Rejected participants (see Section 2.2 for an overview) are not discussed further.

2.5 – Statistical analysis

2.5.1 – Generic ANOVA procedure

All statistical analyses followed the same order of testing. First, mixed-model ANOVAs were used to test for main effects and interactions, with within-subject factors determined by individual tasks design and between-subject factors of device ($\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) or platform ($\text{Computer}_{\text{Labvanced}}$ vs. $\text{Computer}_{\text{Horizon}}$).

All interactions revealed by the omnibus analyses were followed-up with the appropriate repeated-measures (RM) ANOVAs, collapsing over any factors that did not interact. Should any of these interactions be with a between-subjects factor, follow-up analyses always split at the levels of that factor first (i.e., between-subjects factors were not included in any ANOVAs beyond the omnibus). Interactions at subsequent levels of analyses continued to be explored until the simple main effects of each factor were examined at all levels of the other factors. Significant main effects were then explored with all possible pairwise comparisons. All multi-way mixed- and RM-ANOVAs were family-wise error corrected using a sequential Bonferroni procedure (Cramer *et al.*, 2016), and all repeated-measures main effects and interactions were Greenhouse-Geiser corrected to protect against violations of sphericity. Pairwise comparison tests were corrected using a Bonferroni type adjustment to protect against multiple comparisons. Significance was set at a corrected $p<=0.01$.

2.5.2 – Device comparison

Assessments in which differences between computer-, tablet- and smartphone-based testing were compared are described in two categories: within-task assessments and between-measures assessments.

Within-task assessments were conducted for each dependent measure of interest and examined decision difficulty effects as a function of the task-specific within-subjects factors. The main objective of these assessments was to determine whether task-specific effects (as expected by previous studies, e.g., Faulkenberry *et al.*, 2016; Koop & Johnson, 2013; Dale & Duran, 2011; Maldonado *et al*., 2019) were replicated and whether these effects were consistent despite differences in testing device.

Between-measures assessments examined the relationship between measures of decision difficulty (reaction time, movement time and trajectory curvature), with the objective of determining whether there were correlational effects between measures of decision difficulty (e.g., whether increased curvature is correlated with increased movement time), and whether these relationships were consistent across task and testing device.

The specific analyses conducted to assess within-task differences in decision difficulty (Within-task, Section 2.5.2.1) and the relational differences between measures of decision difficulty (Between-measures, Section 2.5.2.2) are outlined below. Where relevant for omnibus mixed-model ANOVAs, device ($\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) was included as a between-subject factor.

2.5.2.1 – Within-task

2.5.2.1.1 – Numeric-Size congruency

For each of the dependent measures (reaction time, movement time and MAD), the following omnibus analysis was conducted:

1) A 3 (Number Pair: 1 – 2 vs. 2 – 8 vs. 8 – 9) x 2 (Numeric-Physical Congruency: Congruent vs. Incongruent) x 2 (Presentation Side: larger number Left vs larger number Right) x 3 (Device: $\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) mixed-model ANOVA

2.5.2.1.2 – Sentence Verification

For each of the dependent measures (reaction time, movement time and MAD), the following omnibus analysis was conducted:

2) A 2 (Truth Value: True vs. False) x 2 (Polarity: Positive vs. Negative) x 3 (Device: $\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) mixed-model ANOVA

2.5.2.1.3 – Photo Preference

Within-task analyses of the Photo Preference trials examining device differences followed two streams of investigation: 1) Differences between conditions in which at least one photo was characterized as being High in pleasantness (High – High, High – Med, and High – Low; referred to as the High-Chosen analysis), with analyses limited only to those where the High photo was selected and 2) Differences between conditions in which the pleasantness of the paired photos were matched (High – High, Med – Med, and Low – Low; referred to as the Matched-Pair analysis).

For each dependent measure (MAD, movement time, reaction time), the omnibus analyses for each of these investigations were formulated as follows:

3) A 3 (Valence Pairing: High – High vs. High – Med vs. High – Low) x 2 (Presentation Side: Left vs. Right) x 3 (Device: $\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) mixed-model ANOVA

4) A 3 (Valence Pairing: High – High vs. Med – Med vs. Low – Low) x 2 (Reach Direction: Left vs. Right) x 3 (Device: $\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) mixed-model ANOVA

2.5.2.2 – Between-measures

To explore the relationship between measures of decision difficulty, a Pearson's correlation coefficient (*r*) indicating the direction and the strength of the relation between each measure (MAD – movement time, MAD – reaction time, and reaction time – movement time) was obtained for each participant within each condition, task, and device. Differences in between-measure correlations between task and device were then assessed using an omnibus analysis formulated as follows:

5) A 3 (Correlated Coefficients: $r_{\text{MAD,MT}}$ vs. $r_{\text{MAD,RT}}$ vs. $r_{\text{MT,RT}}$) x 3 (Task: Numeric-Size Congruency vs. Sentence Verification vs. Photo Preference) x 3 (Device: $\text{Computer}_{\text{Labvanced}}$ vs. $\text{Tablet}_{\text{Labvanced}}$ vs. $\text{Smartphone}_{\text{Labvanced}}$) mixed-model ANOVA

2.5.3 – Platform comparison

Analyses exploring differences between Labvanced and Horizon platforms, including within-task and between-measure assessments, were identical to those assessing differences between devices, apart from platform ($\text{Computer}_{\text{Labvanced}}$ vs. $\text{Computer}_{\text{Horizon}}$) being included as a between-subject factor when relevant.

2.6 – Predictions

2.6.1 – Device comparison

2.6.1.1 – Within-task

Movement time, reaction time, and trajectory curvature (operationalized as MAD, see Section 2.4.1.3) were measured as indexes of decision difficulty. Broadly, we expected these measures of decision difficulty in each decision domain task to illustrate the task-specific effects described in Section 1.6.2.

In the Numeric-Size Congruency task, we predicted we would replicate classic congruency effect results previously demonstrated by Faulkenberry and colleagues (2016) wherein incongruent trials display greater relative decision difficulty, with greater congruency-dependent effects seen for number pairs that are further apart numerically (as exemplified in Figure 1.3). Moreover, we believed increased decision difficulty would be reflected as a relative increase in scores across all three measures of interest. That is, incongruent trials would display greater reaction times, greater movement times and greater trajectory curvature, with exaggerated effects for 2 – 8 number pairs compared to 1 – 2 and 8 – 9 number pairs (see Section 1.6.2.1). We did not predict any differences between 1 – 2 and 8 – 9 number pairs within the same congruency condition.

In the Sentence Verification task, we expected to replicate classic truth value-negation interaction effects previously demonstrated by Dale and Duran (2010) and later by Maldonado and colleagues (2019; see Section 1.6.2.2). Specifically, we predicted greater reaction times, movement times and mouse trajectory curvature – each indicating increased relative decision difficulty – for statements containing a negation (TN and FN) compared to those not containing a negation (TP and FP), with greater increases in decision difficulty when true sentences are negated (TN) compared to false sentences (FN; as exemplified in Figure 1.4).

In the Photo Preference task, we expected to find an ordinal increase in decision difficulty as a function of photo pleasantness similarity as was previously shown by Koop and Johnson (2013, exemplified in Figure 1.5; see Section 1.6.2.3). Specifically, we believed the High-Chosen analysis (in which High – High, High – Med, and High – Low photo pairings are compared) would show a linear increase in relative difficulty as the High photo counter-pair increased in pleasantness. As such, similar to Koop and Johnson (2013), we believed we would find greater reaction times, movement times and trajectory curvature for High – High trials, followed by High – Med trials and finally High – Low trials. In the Matched-Pair analysis (comparing High – High, Med – Med, and Low – Low photo pairings) we expected no difference between photo pairs to be reflected by the selected measures of decision difficulty as there are no disparities in pleasantness between pairs to drive differences in relative difficulty.

Overall, we believed each mouse-tracked measure of difficulty to accurately express the evidence accumulation processes underlying task decisions (Sullivan *et al.,* 2015; Stillman *et al*., 2020), with more difficult decisions the result of greater competitive pull between choice options. We believed this choice competition would first be reflected in greater reaction times (the result of more time spent accumulating and processing evidence prior to movement), and finally give rise to greater trajectory curvature towards the unchosen option (a physical manifestation of choice competition) and movement times.

Regarding our examination of measure consistency across device (computer, tablet, smartphone), we predicted that the expression of decision difficulty in each dependent measure (trajectory curvature, movement time, reaction time) would vary as a result of the device used for testing. As studies have shown similar patterns of reach-trajectory results between 3-D reaches, mouse movements and stylus movements (Moher & Song, 2019), we did not anticipate differences as a result of the user-interaction profiles of the devices tested (e.g., mouse movement vs. touchscreen finger swipe). Instead, we believed the sensitivity to post-reaction time measures of decision difficulty (trajectory curvature and movement time) would be reliant on the size of the device in use, with smaller devices providing less surface area over which decision difficulty could be expressed. Ultimately, we believed constraining trajectories to a smaller surface area would allow for less variability in movement between task choices to be expressed, consequently producing smaller differences in trajectories between trials previously exhibiting different degrees of decision difficulty. An overall reduction in movement time was also expected as the spatial distance between the start position and choice options is reduced (resulting in more ballistic movements) and the complexity of trajectory curvature is limited (requiring less time to perform the movement and resulting in less variability in movement times). As such, we hypothesized that trajectory curvature and movement time would be most sensitive to decision difficulty during computer use, with a reduction in sensitivity during tablet- and finally smartphone-use. We did not, however, expect device-dependent changes in sensitivity within reaction time measures.

2.6.1.2 – Between-measures

Predictions regarding the relationship between measures of decision difficulty (trajectory curvature, movement time and reaction time) were primarily founded on their theorized role in the expression of decision processes within an evidence accumulation model of decision making (Ratcliff & McKoon, 2008; Wispinski *et al.*, 2020; Stillman *et al.,* 2020). First, reaction time is thought to reflect decision processing time prior to movement onset, which fluctuates with the need for further evidence prior to reaching a decision threshold. As more time is required to accumulate sufficient evidence in support of one choice over another when competition between choice is high (i.e., it is a difficult decision), reaction time is thought to provide a direct measure of decision difficulty. As these processes have been shown to leak beyond reaction time to be reflected in movement, trajectory curvature also provides a direct measure of decision difficulty, with increases in curvature reflecting increased competitive pull between options. Finally, rather than independently providing a read-out of decision difficulty, movement times are thought to arise as a composite of movement vigor (reflected in the velocity of the movement and dependent on confidence; Dotan, *et al.*, 2019) and movement path (degree of trajectory curvature).

Independent of whether a measure is thought to provide a direct or indirect quantification of relative decision difficulty, each has been shown to increase (greater trajectory curvature, reaction times and movement times) as decision difficulty increases. As such, we predicted that each of the three measures of decision difficulty (MAD, movement time and reaction time) would be positively correlated, such that trials with increased trajectory curvature would also demonstrate increase movement and reaction times. Further, we believed this relationship would remain consistent across all three decision domain tasks.

We also hypothesized that the strength of the relationship between post-reaction time measures (trajectory curvature and movement time) and reaction time would vary according to the testing device. Specifically, we believed that the correlations between trajectory curvature and reaction time (MAD – RT) and movement time and reaction time (MT – RT) would be most positively correlated during computer use, followed by tablet use and finally smartphone use. This hypothesis is founded on the earlier prediction that sensitivity to trajectory curvature differences, and thus movement time differences, will become diminished as the testing device decreases in size but reaction time will remain a consistently sensitive measure across devices.

2.6.2 – Platform comparison

2.6.2.1 – Within-task

Despite the differences in data processing and cleaning strategies between the Labvanced and Horizon platforms, we did not predict any differences between platforms. As this analysis was conducted on data derived exclusively from computer use (mitigating any earlier device-specific predictions), we expected measures of decision difficulty captured by both platforms to illustrate the task-specific effects described in Section 1.6.2 (see Section 2.6.1.1 for an overview of these predictions), with no significant differences between $\text{Computer}_{\text{Labvanced}}$ and $\text{Computer}_{\text{Horizon}}$ factors.

2.6.2.2 – Between-measures

As with the device comparison analysis, we again hypothesized that the measures of reach behaviour in each task would be mirrored by the temporal measures of performance. Specifically, trials exhibiting the greatest trajectory curvature would also demonstrate longer movement and reaction times, and all measures would be significantly positively correlated independent of the task from which they were derived. As both platforms were tested on a singular device in this analysis, we did not predict any differences in the relationship between measures as a function of platform.

3.0 – Results

3.1 – Device comparison

3.1.1 – Within-task results

3.1.1.1 – Numeric-Size Congruency

Discussed previously (Section 2.6.1.1), for the Numeric-Size Congruency task we predicted a successful replication of results reported by Faulkenberry and colleagues (2016) wherein trials presenting digit options Incongruent in numeric and physical size (e.g., 2 vs. 8) reflected increased decision difficulty in each of the three collected measures of interest – reaction time, movement time and trajectory curvature (MAD) – relative to Congruent trials, with greater congruency effects for pairs of greater numerical distance (e.g., 2 – 8) compared to those closer in numerical value (e.g., 1 – 2 and 8 – 9). Given our experimental task design, this projected numeric distance-modulated congruency effect would be reflected in an interaction between Congruency and Number Pairs factors. Further, should this effect be susceptible to changes in device (as previously predicted for movement time and trajectory curvature measures, see Section 2.6.1.1), we would expect to see a three-way interaction between Congruency, Number Pair and Device factors (or a main effect of Device at minimum should any variability in response be due to Device differences).

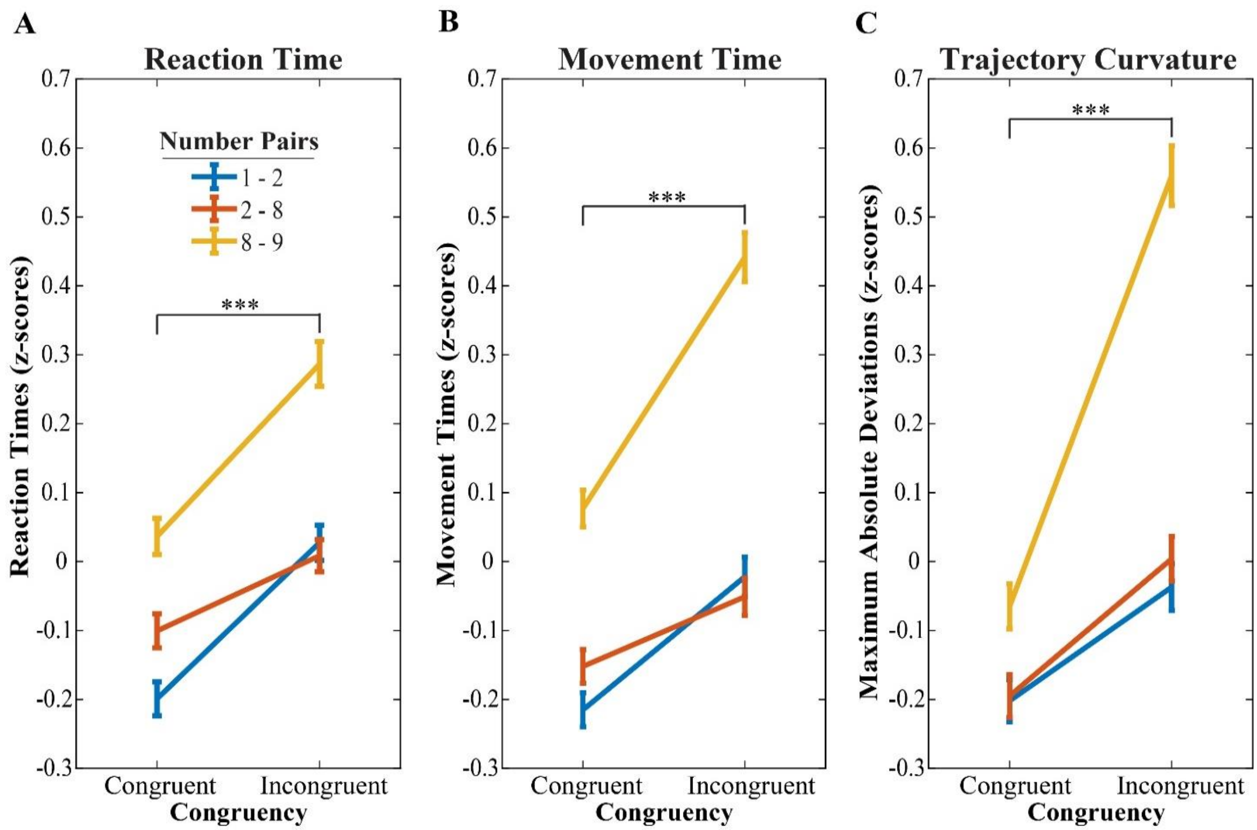

Figure 3.1 Numeric-Size Congruency mean standardized scores for **A)** Reaction time, **B)** Movement time and **C)** Trajectory curvature measures, demonstrating the interaction between Number Pairs and Congruency factors. Incongruent trials were found to be significantly different from Congruent trials at all levels of Number Pairs, and across all three measures. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

3.1.1.1.1 – Reaction time analysis

For Numeric-Size Congruency reaction time data, the 4-factor Number Pairs x Congruency x Presentation Side x Device mixed-model ANOVA revealed main effects of Number Pairs ($F_{(2,472)} = 76.43$, $p = 4.18\text{e-}27$, $\eta^2 = 0.060$), Congruency ($F_{(1,236)} = 137.12$, $p = 2.85\text{e-}25$, $\eta^2 = 0.050$), and Device ($F_{(1,236)} = 12.36$, $p = 7.83\text{e-}6$, $\eta^2 = 3.08\text{e-}4$), as well as a two-way Number Pairs x Congruency interaction ($F_{(2,472)} = 8.77$, $p = 1.82\text{e-}4$, $\eta^2 = 0.005$) and a two-way Number Pairs x Device interaction ($F_{(4,472)} = 14.13$, $p = 3.96\text{e-}10$, $\eta^2 = 0.022$), both of which were followed up by separate simple main effect 1-factor RM-ANOVAs.

Simple main effect 1-factor RM-ANOVAs assessing Congruency at each level of Number Pairs revealed significant main effects of Congruency at all three levels (1 – 2: $F_{(1)} = 77.74$, p = 2.64e-16; 2 – 8: $F_{(1)} = 18.15$, p = 2.95e-5; 8 – 9: $F_{(1)} = 75.93$, p = 5.27e-16). As predicted, Incongruent pairings showed greater standardized reaction time scores compared to Congruent pairings at all three levels, indicating significant difference in decision difficulty as a function of numeric-size Congruency wherein Incongruent trials are more difficult than Congruent trials. Interestingly, however, while these effects are reliably modulated by the numeric values of the digits being compared, rather than pairs with the greatest numeric distance (e.g., 2 – 8) displaying greatest overall difficulty as predicted, 8 – 9 digit pairings displayed the greatest overall decision difficulty and greatest congruency effects (see Table A.3.1 for mean scores). This interaction is illustrated in Figure 3.1 A, and discussed further below (Section 4.1). Of note, no three-way Number Pairs x Congruency x Device interaction was uncovered, suggesting that, as predicted, decision difficulty expressed through measures of reaction time did not differ between devices.

To explore the interaction between Number Pairs and Device, three simple main effect 1-factor RM-ANOVAs were used to test Number Pairs differences at each level of Device. A main effect of Number Pair was revealed at each of the levels of Device (Computer: $F_{(2)} = 78.30$, p = 1.35e-24; Tablet: $F_{(2)} = 7.21$, $p = 0.001$; Smartphone: $F_{(2)} = 13.73$, p = 3.27e-6). Pairwise comparisons showed no significant difference in standardized reaction times when a Tablet was used as a testing device. For Computer-acquired reaction times, pairwise comparisons revealed a significant difference between 1 – 2 and 2 – 8($p = 0.007$), 1 – 2 and 8 – 9 (p = 3.06e-31), and 2 – 8 and 8 – 9 (p = 7.46e-17), giving rise to a linear relationship in which 8 – 9 pairings showed greater reaction times than 2 – 8 pairings, which in turn had greater reaction times than 1 – 2 pairings. Pairwise comparisons for Smartphone-acquired reaction times revealed only significant differences between 1 – 2 and 8 – 9 pairings (p = 4.60e-4), and 2 – 8 and 8 – 9 pairings (p = 8.64e-4), with 8 – 9 pairings showing greater reaction times in both instances. Taken together, these results suggest a difference in decision difficulty between number pair comparisons wherein 8 – 9 pairings are most difficult (see Figure 3.2 A and Table A.3.1 for mean scores). Of note, however, these differences are only captured through Computer and Smartphone use. Though not predicted, speculations over the source of these results are discussed in Section 4.2.

3.1.1.1.2 – Movement time analysis

For Size Congruency movement time data, the 4-factor Number Pairs x Congruency x Presentation Side x Device mixed-model ANOVA revealed main effects of Number Pairs ($F_{(2,472)}$ = 233.51, p = 9.81e-69, η^2 = 0.127), Congruency ($F_{(1,236)}$ = 183.99, p = 2.26e-31, η^2 = 0.051), and Presentation Side ($F_{(1,236)}$ = 60.52, p = 2.27e-13, η^2 = 0.033), as well as a two-way Number Pairs x Congruency interaction ($F_{(2,472)}$ = 22.17, p = 9.52e-10, η^2 = 0.012), a two-way Congruency x Device interaction ($F_{(2,236)}$ = 16.33, p = 2.28e-7, η^2 = 0.009) and a two-way Presentation Side X Device interaction ($F_{(2,236)}$ = 17.78, p = 6.40e-8, η^2 = 0.019).

Subsequent simple main effect 1-factor RM-ANOVAs assessing Congruency at each level of Number Pairs revealed significant main effects of Congruency at all three levels (1 – 2: $F_{(1)}$ = 47050, p = 4.97e-11; 2 – 8: $F_{(1)}$ = 17.32, p = 4.44e-5; 8 – 9: $F_{(1)}$ = 134.80, p = 5.98e-25). As with the reaction time scores, Incongruent pairs showed greater standardized movement time scores compared to Congruent pairs in all cases (see Table A.3.1 for mean scores). Depicted in Figure 3.1 B, this interaction again suggests greater relative decision difficulty for Incongruent trials, with 8 – 9 pairs showing greater overall difficulty (i.e., longest movement times) and congruency effects (see Section 4.1 for discussion). Of note, while we predicted that decision difficulty effects expressed in movement time would diminish as device size shrank, the absence of an interaction between these two interacting factors and Device suggest instead that this effect does not differ across the computer-, tablet-, and smartphone-based testing.

To explore the interaction between Congruency and Device, three simple main effect 1-factor ANOVAs were used to test Congruency at each level of Device. All three levels of Device showed a significant main effects of Congruency (Computer: $F_{(1)} = 11.19$, $p = 8.99e\text{-}4$; Tablet: $F_{(1)} = 84.34$, $p = 5.31e\text{-}14$; Smartphone: $F_{(1)} = 111.39$, $p = 1.29e\text{-}16$), with Incongruent pairs showing greater standardized movement time scores compared to Congruent pairs (see Table A.3.1 for mean scores). Discussed in Section 4.2 and illustrated in Figure 3.2 B, it appears that this interaction is driven by diminished significance of congruency effects within Computer-acquired data compared to Tablet- and Smartphone-acquired data.

Finally, three simple main effect 1-factor RM-ANOVAs were used to explore the interaction between Presentation Side and Device, testing Presentation Side at each level of Device. A main effect of Presentation Side was revealed for Tablet- ($F_{(1)} = 48.58$, $p = 9.48e\text{-}10$) and Smartphone- ($F_{(1)} = 35.20$, $p = 7.99e\text{-}8$) acquired standardized movement scores, but not those acquired through Computer use. For both Tablet and Smartphone device use, greater standardized movement time scores occurred when the numerically larger number was presented on the right (mandating a rightward reach) compared to when the numerically larger number was presented on the left (mandating a leftward reach; see Table A.3.1 for mean scores and Section 4.1 for discussion).

3.1.1.1.3 – Trajectory analysis

For Size Congruency MAD trajectory data, the 4-factor Number Pairs x Congruency x Presentation Side x Device mixed-model ANOVA revealed main effects of Number Pairs ($F_{(2,472)}$

= 232.41, p = 3.95e-66, η^2 = 0.076), Congruency ($F_{(1,236)}$ = 391.60, p = 4.94e-52, η^2 = 0.074), and Presentation Side ($F_{(1,236)}$ = 43.65, p = 2.59e-10, η^2 = 0.063), as well as a two-way Number Pairs x Congruency interaction ($F_{(2,472)}$ = 98.25, p = 2.09e-33, η^2 = 0.030) and a two-way Presentation Side x Device interaction ($F_{(2,236)}$ = 17.09, p = 1.17e-7, η^2 = 0.049).

As with reaction time and movement time analyses, three simple main effect 1-factor RM-ANOVAs were used to compare Congruency at each level of Number Pairs, and revealed significant main effects of Congruency at all three levels (1 – 2: $F_{(1)}$ = 54.88, p = 2.26e-12; 2 – 8: $F_{(1)}$ = 80.75, p = 8.42e-17; 8 – 9: $F_{(1)}$ = 331.17, p = 7.05e-47). In all cases, Incongruent pairs again showed greater standardized MAD scores compared to Congruent pairs (see Table A.3.1 for mean scores), with greatest congruency effect revealed for 8 – 9 pairs compared to 1 – 2 and 2 – 8 pairs. Illustrated in Figure 3.1 C, these results align with the decision difficulty effects revealed within reaction time and movement time, ultimately replicating predicted results with the exception of a numeric distance effect driving the interaction (see Section 4.1 for discussion). Contrary to our predictions, however, no interaction between our effect of interest (numeric distanced-modulated congruency, as would be reflected in a Number Pair x Congruency x Device interaction) and Device was revealed. As with our movement time analysis, this again suggests that these effects do not differ across the computer-, tablet-, and smartphone-based testing.

To explore the Presentation Side x Device interaction, three simple main effect 1-factor RM-ANOVAs tested Presentation Side at each level of Device. A main effect of Presentation Side was revealed for Tablet- ($F_{(1)} = 21.67$, $p = 1.33e-5$) and Smartphone- ($F_{(1)} = 33.33$, $p = 1.56e-7$) acquired standardized MAD scores, but not those acquired through Computer use (Figure 3.2 C). For both Tablet and Smartphone device use, greater standardized MAD scores occurred when the numerically larger number was presented on the right (mandating a rightward reach) compared to when the numerically larger number was presented on the left (mandating a leftward reach; see Table A.3.1 for mean scores). Speculations as to the reason for these differences are discussed in Section 4.2.

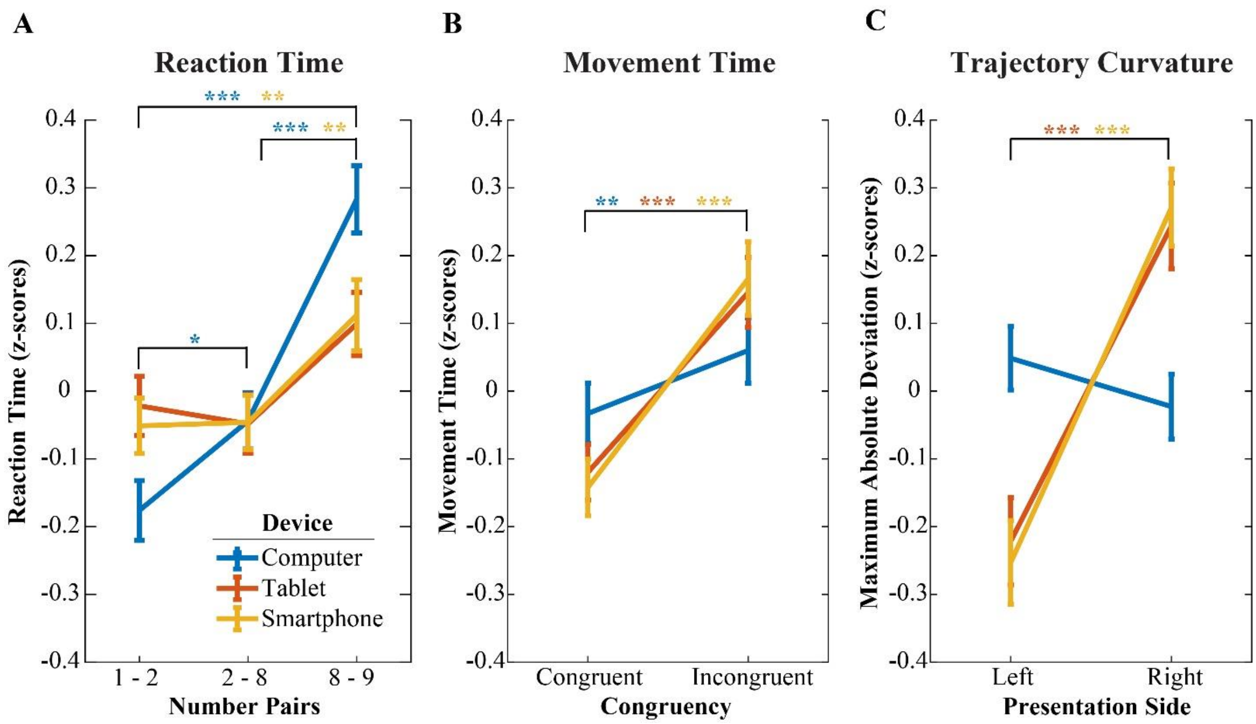

Figure 3.2 Numeric-Size Congruency mean standardized scores for **A)** Reaction time, **B)** Movement time and **C)** Trajectory curvature measures, each demonstrating factor interactions with Device. Significant differences between factor levels are indicated in the colour corresponding to the level of Device in which they were revealed. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

3.1.1.2 – Sentence Verification

Discussed in Section 2.6.1.1, for the Sentence Verification task we predicted a replication of results reported by Dale and Duran (2010) and Maldonado and colleagues (2019) in which trials presenting negated statements (TN and FN) reflect increased decision difficulty in each of the three collected measures of interest – reaction time, movement time and trajectory curvature (MAD) – relative non-negated statements (TP and FP), with greater negation-driven effects for true statements compared to false statements. Given our task design, this effect would be reflected in a Truth Value x Polarity interaction. Further, should this effect be susceptible to changes in device (as previously predicted for movement time and trajectory curvature measures, see Section 2.6.1.1), we would also expect to see these factors interact with device (giving rise to a three-way Truth Value x Polarity x Device interaction).

3.1.1.2.1 – Reaction time analysis

The 3-factor Truth Value x Polarity x Device mixed-model ANOVA for Sentence Verification standardized reaction time scores revealed main effects of Polarity ($F_{(1,236)}$ = 1011.26, p = 2.76e-87, η^2 = 0.621), a two-way Truth Value x Polarity interaction ($F_{(1,236)}$ = 245.32, p = 2.17e-38, η^2 = 0.085), a two-way Polarity x Device interaction ($F_{(2,236)}$ = 10.19, p = 5.69e-5, η^2 = 0.013) and a three-way Truth x Polarity x Device interaction ($F_{(2,236)}$ = 7.99, p = 4.39e-4, η^2 = 0.005). To investigate the three-way interaction further, a 2-factor Truth x Polarity RM-ANOVA was conducted at each level of Device.

Analyses at each of the three levels of Device revealed a main effect of Polarity (Computer: $F_{(1,82)} = 670.46$, $p = 3.15\text{e-}41$, $\eta^2 = 0.702$; Tablet: $F_{(1,77)} = 241.98$, $p = 1.78\text{e-}25$, $\eta^2 = 0.592$; Smartphone: $F_{(1,77)} = 231.63$, $p = 6.23\text{e-}25$, $\eta^2 = 0.583$) and a two-way Truth Value x Polarity interaction (Computer: $F_{(1,82)} = 191.58$, $p = 3.67\text{e-}23$, $\eta^2 = 0.106$; Tablet: $F_{(1,77)} = 53.08$, $p = 2.38\text{e-}10$, $\eta^2 = 0.054$; Smartphone: $F_{(1,77)} = 46.36$, $p = 1.91\text{e-}09$, $\eta^2 = 0.060$), but no main effect of Truth Value. At each level of Device, follow-up simple main effect 1-factor RM-ANOVAs examining Polarity at each level of Truth Value also revealed a main effect of Polarity both True statements (Computer: $F_{(1)} = 706.09$, $p = 4.72\text{e-}42$; Tablet: $F_{(1)} = 211.12$, $p = 9.12\text{e-}24$; Smartphone: $F_{(1)} = 177.66$, $p = 1.08\text{e-}21$) and False statements (Computer: $F_{(1)} = 205.34$, $p =$

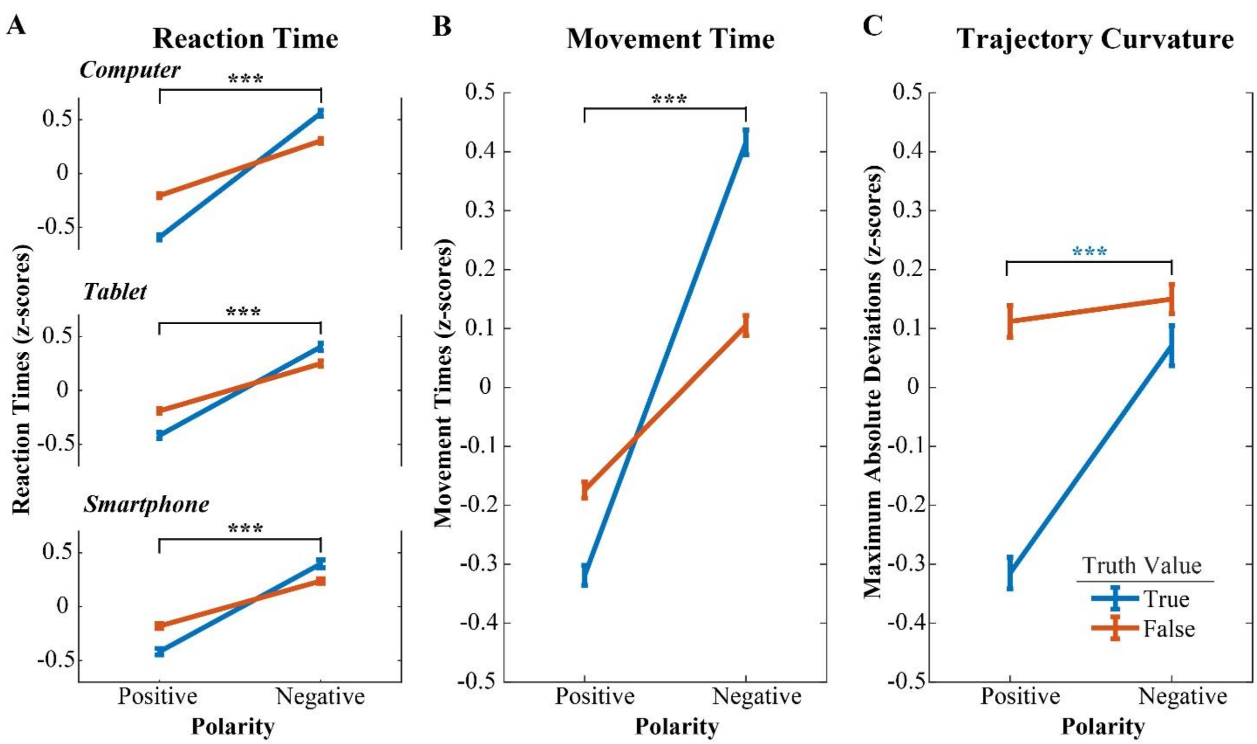

Figure 3.3 Sentence Verification mean standardized scores for **A)** Reaction time, **B)** Movement time and **C)** Trajectory curvature measures, illustrating the interaction between Truth Value and Polarity. Significant differences between Positive and Negative polarities are indicated in black if applicable to all levels of Truth Value within that measure, or, when appropriate, in the colour corresponding to the level of Truth Value in which they were revealed. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

4.86e-24; Tablet: $F_{(1)} = 131.82$, $p = 2.36e\text{-}18$; Smartphone: $F_{(1,)} = 140.94$, $p = 4.50e\text{-}19$). At all levels of Device and Truth value, the main effect of Polarity was driven by Negative statements exhibiting greater standardized reaction time scores than Positive statements (see Table A.3.2 for mean scores). Taken together, these results suggest a replication of predicted negation and truth-value driven decision difficulty effects across all three devices, with device differences in reaction time driven only by differences in effect significance. The interaction between Truth Value and Polarity, is illustrated in Figure 3.3 A and 3.4 A.

3.1.1.2.2 – Movement time analysis

The 3-factor Truth Value x Polarity x Device mixed-model ANOVA for Sentence Verification standardized movement time scores revealed main effects for Truth Value ($F_{(1,236)} = 16.80$, $p = 5.72e\text{-}5$, $\eta^2 = 0.012$), Polarity ($F_{(1,236)} = 609.75$, $p = 2.367e\text{-}67$, $\eta^2 = 0.425$), and a two-way interaction between these factors ($F_{(1,236)} = 189.64$, $p = 4.63e\text{-}32$, $\eta^2 = 0.085$; see Figure 3.3 B for an illustration of this interaction). A two-way Polarity x Device interaction was also revealed ($F_{(2,236)} = 19.42$, $p = 1.56e\text{-}8$, $\eta^2 = 0.027$; see Figure 3.4 B). Each interaction was followed up by separate simple main effect 1-factor RM-ANOVAs examining Polarity at each level of the other factor.

Tests conducted to follow up the Truth Value x Polarity interaction revealed a main effect of Polarity when statements were True ($F_{(1)} = 614.95$, $p = 1.15e\text{-}67$) as well as when they were False ($F_{(1)} = 153.07$, $p = 1.97e\text{-}27$). In both cases, Negative statement showed greater standardized movement time scores compared to Positive statements (see Table A.3.2 for score means). Illustrated in Figure 3.3 B, this particular result supports a replication of previous results (Dale & Duran, 2010; Maldonado *et al.*, 2019). Specifically, as with reaction time, we see an increase in decision difficulty expressed by movement time during the verification of negated

(Negative) statements compared to the verification of non-negated statements (Positive), with greater differences between negation conditions for True statements than False statements. Contrary to prior predictions, however, these interconnected truth value and negation effects do not interact with device, suggesting consistent replication of these effects across computer-, tablet- and smartphone-based testing.

Upon examination of the Polarity x Device interaction, a main effect of Polarity was also revealed at each of the three levels of Device (Computer: $F_{(1)} = 79.32$, $p = 1.11e\text{-}13$; Tablet: $F_{(1)} = 202.45$, $p = 2.97e\text{-}23$; Smartphone: $F_{(1)} = 472.56$, $p = 1.34e\text{-}34$). In all three cases, Negative statements showed greater standardized movement time scores compared to Positive statements (Figure 3.4 B, see Table A.3.2 for mean scores). While tangential to the main effects of interest for this task, differences in movement time-reflected negation effects are discussed with respect to device differences in Section 4.2.

3.1.1.2.3 – Trajectory analysis

The 3-factor Truth Value x Polarity x Device mixed-model ANOVA for Sentence Verification standardized MAD scores revealed main effects for Truth Value ($F_{(1,236)} = 30.22$, $p = 9.97e\text{-}8$, $\eta^2 = 0.074$), Polarity ($F_{(1,236)} = 123.69$, $p = 2.22e\text{-}23$, $\eta^2 = 0.050$), and a two-way interaction between these factors ($F_{(1,236)} = 95.23$, $p = 4.07e\text{-}19$, $\eta^2 = 0.033$; see Figure 3.3 C for an illustration of this interaction). A two-way Truth x Device interaction was also revealed ($F_{(2,236)} = 14.99$, $p = 7.39e\text{-}7$, $\eta^2 = 0.074$).

To further understand the interaction between Truth Value and Polarity, a simple main effect 1-factor RM-ANOVA was conducted to assess differences in Polarity at each level of Truth Value. A main effect of Polarity was revealed when statements were True ($F_{(1)} = 154.28$, $p = 1.36e\text{-}27$), but not when they were False (see Figure 3.3 C). For True statements, statements that were Negative in polarity showed greater standardized MAD values compared to those that were Positive (see Table A.3.2 for mean scores). As with reaction time and movement time results, these results suggest negation-driven decision difficulty effects which are diminished when statements are False compared to when they are True. Device was not shown to modulate this interaction as it is reflected in trajectory curvature.

To explore the interaction between Truth Value and Device, a subsequent simple main effect 1-factor RM-ANOVA was used to explore difference in Truth Value at each level of Device. A main effect of Truth Value was revealed at the level of Tablet ($F_{(1)} = 14.88$, $p = 2.36e\text{-}4$) and Smartphone ($F_{(1)} = 29.38$, $p = 6.59e\text{-}7$), but not Computer. For both Tablet and Smartphone, False statements showed greater standardized MAD scores compared to True statements (Figure 3.3 C, see Table A.3.2 for mean scores). Again suggesting that possible nuanced differences may

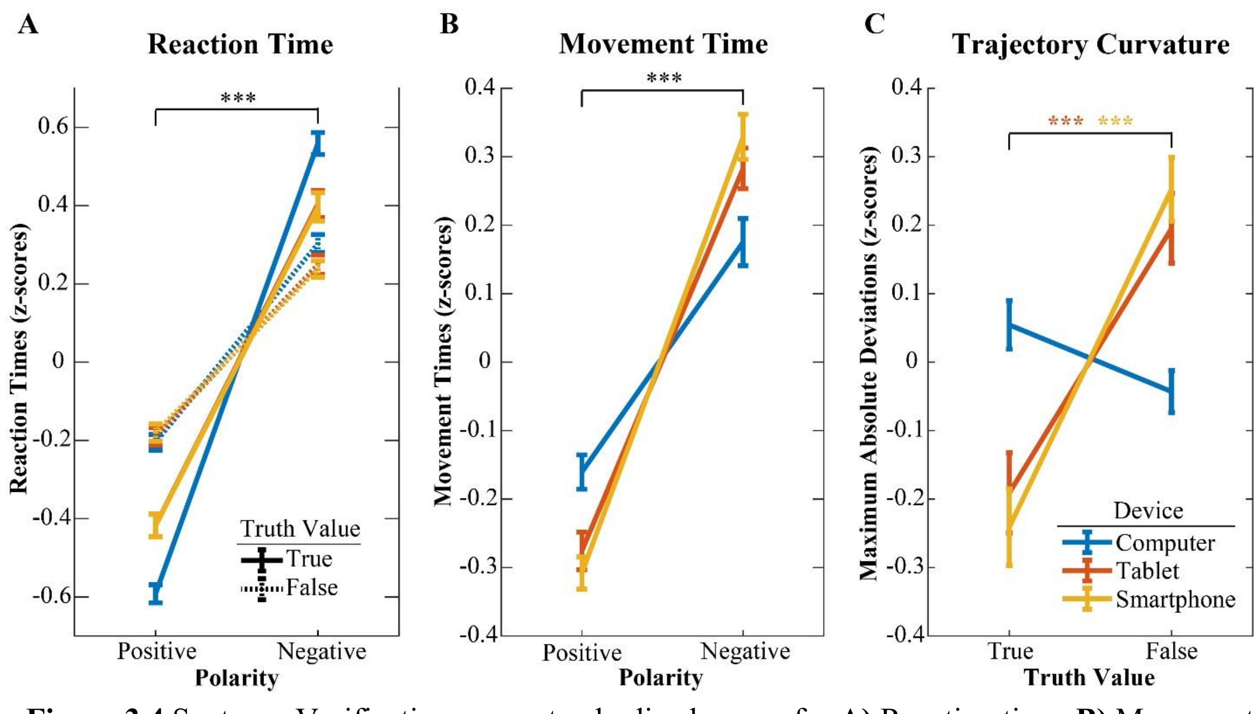

Figure 3.4 Sentence Verification mean standardized scores for **A)** Reaction time, **B)** Movement time and **C)** Trajectory curvature measures, each demonstrating factor interactions with Device. Significant differences between factor levels are indicated in the colour corresponding to the level of Device in which they were revealed or in black should they consistent across all levels. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

exist between decision difficulty expressed during different device use, though tangential to our replication of task-specific effects, these results are further discussed in Section 4.2.

3.1.1.3 – Photo Preference

Photo choice selections in the Photo Preference task revealed a global preference for photos rated as more pleasant ($M_{\text{More Pleasant Selected}}$ = 78.3%), substantiating claims that preference is roughly analogous with pleasantness ratings (Koop & Johnson, 2013). For Photo Preference trials in which the photo pair options were matched in pleasantness, photos on the left were selected 48.6%, with no difference from 50% (equal chance of selecting left vs. right) for Med – Med trials, a slight leftward bias for High – High pairings (p = 0.0005, P(Left Chosen) = 0.55) and a slight rightward bias for Low – Low pairings (p = 4.88e-14, P(Left chosen) = 0.40).

Subsequent Photo Preference analyses were then separated into two categories: High-Chosen analyses in which only trials containing a High pleasantness photo and in which the High photo was selected were analyzed, and Matched-Pair analyses in which only trials presenting photos matched in pleasantness were analyzed. Based on this segregated analysis, our predictions were twofold: 1) For High-Chosen trials, pairs more similar in photo preference (e.g., High – High) would demonstrate greater measures of decision difficulty than those less similar (e.g., High – Low), with a linear increase as similarity increase, and 2) For Matched-Pair trials, no differences would be revealed between photo pairs.

3.1.1.3.2 – High-Chosen analysis

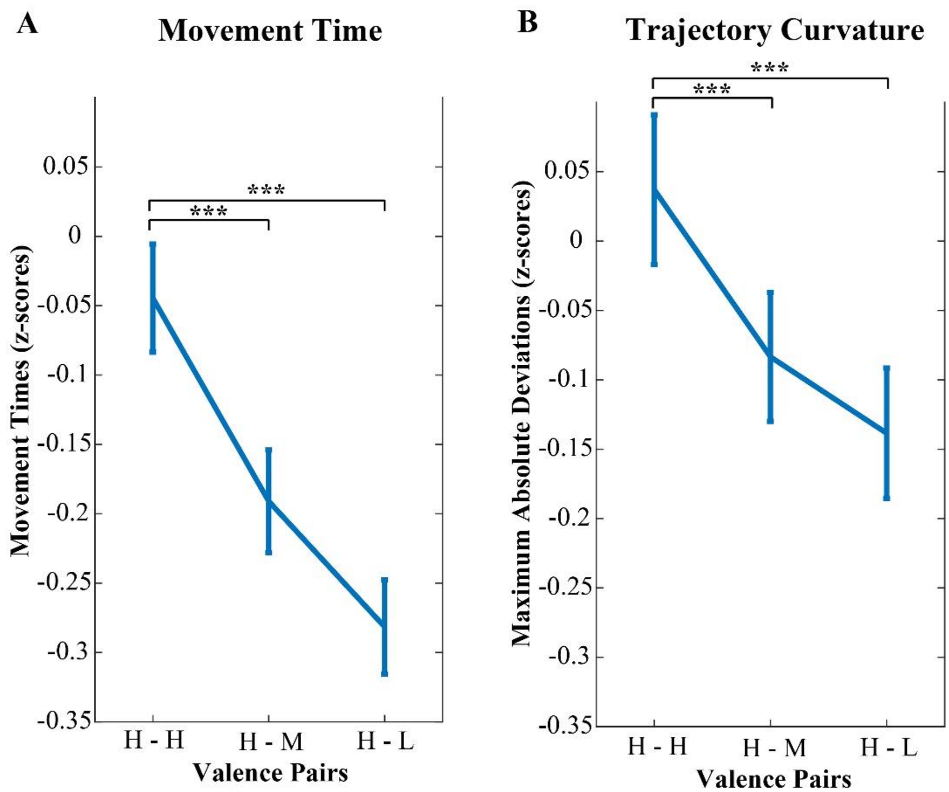

Figure 3.5 Photo Preference High-Chosen mean standardized scores for **A)** Movement time and **B)** Trajectory curvature measures, illustrating the main effect of Valence Pairs. Significant differences between High – High (H – H), High – Med (H – M) and High – Low (H – L) levels of Valence Pairs are indicated where relevant. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

3.1.1.3.2.1 – Reaction time analysis

A 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to Photo Preference High-Chosen standardized reaction time scores revealed no main effects or interactions, suggesting no significant difference between trial conditions or devices used.

3.1.1.3.2.2 – Movement time analysis

A 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to Photo Preference High-Chosen standardized movement time scores revealed only a main effect of Valence Pairing ($F_{(1,232)} = 22.97$, $p = 2.16\text{e-}9$, $\eta^2 = 0.056$), with pairwise comparisons showing a significant difference between High – High and High – Med pairings (p = 1.69e-04), and High – High and High – Low pairings (p = 4.11e-10), but not High – Med and High – Low pairings. Despite the lack of significant difference between High – Med and High – Low pairings, averaged participant means of each condition reveal a linear trend in which High – High trials show the greatest standardized movement time values, followed by High – Med trials and finally High – Low trials (see Figure 3.5 A, see Table A.3.3 for mean scores). Paired with the absence of any interaction with Device, these results suggest that decision difficulty does indeed increase as similarity in photo pleasantness increases and that this effect does not differ as a result of testing device.

3.1.1.3.2.3 – Trajectory analysis

As with the movement time analysis, the 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to Photo Preference High-Chosen standardized MAD scores revealed only a main effect of Valence Pairing ($F_{(1,232)} = 10.92$, p = 4.51e-5, $\eta^2 = 0.018$). Pairwise comparisons showed a significant difference between High – High and High – Med pairings (p = 0.009), and High – High and High – Low pairings (p = 2.05e-05), but not High – Med and High – Low pairings. Mirroring movement times, although not all significantly different, averaged participant standardized MAD score means again showed a linear trend in which High – High trials show the greatest standardized movement times, followed by High – Med trials and finally High – Low trials (see Figure 3.5 B; see Table A.3.3 for mean scores). These results again support our prediction that decision difficulty increased as photo pleasantness increased in similarity. Contrary to our predictions, however, this analysis again did not reveal any interaction or main effect of Device, serving as an additional indicator that trajectory-curvature expressed decision difficulty is consistent independent of testing device used.

3.1.1.3.3 – Matched Pair analysis

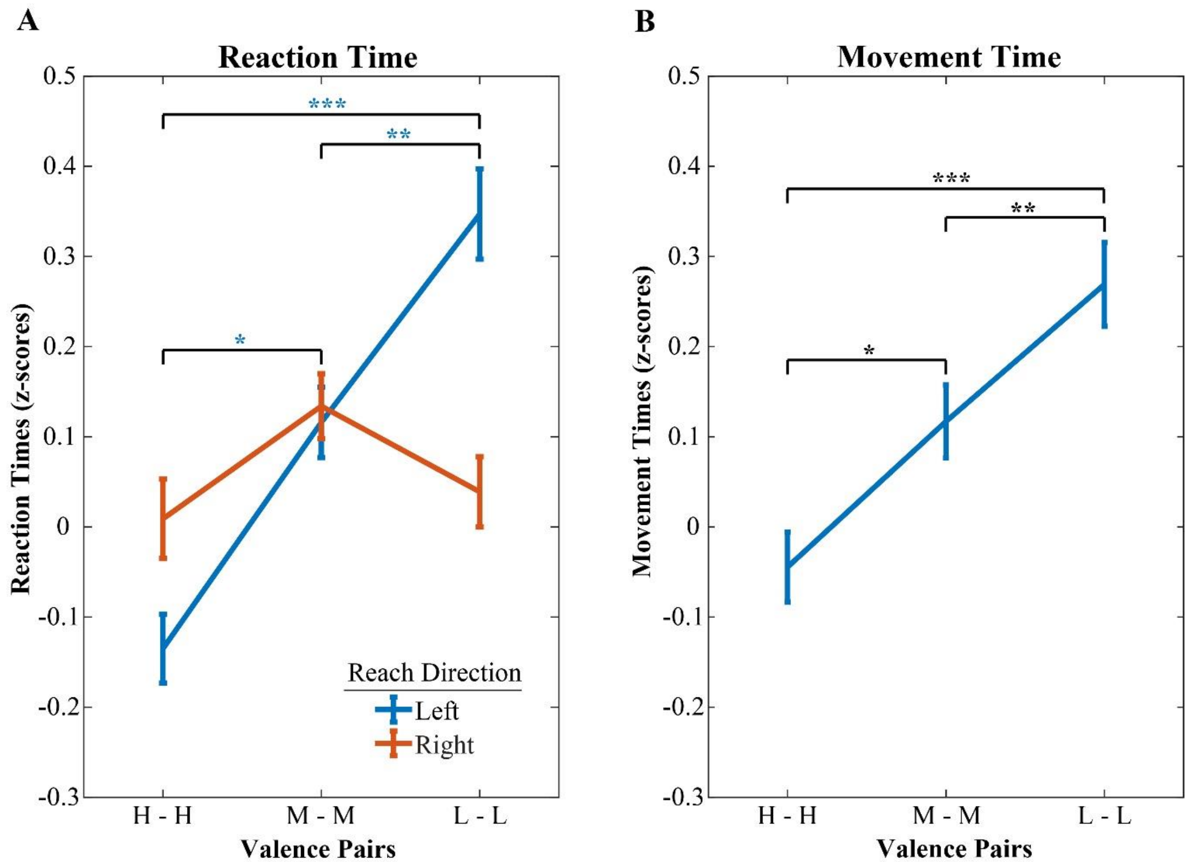

Figure 3.6 Photo Preference Matched-Pair mean standardized scores for **A)** Reaction time and **B)** Movement time measures, illustrating the interaction between Valence Pairs and Reach Direction within Reaction time and the main effect of Valence Pairs within Movement time. Significant differences between High – High (H – H), Med – Med (M – M) and Low – Low (L – L) levels of Valence Pairs are indicated where relevant, with significance reported in the colour corresponding to the level of Reach Direction in which they were revealed for Reaction time. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

3.1.1.3.3.1 – Reaction time analysis

A 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to the Photo Preference Matched-Pair standardized reaction time scores revealed a main effect of Valence Pair ($F_{(1,234)} = 17.37$, $p = 1.06e{-7}$, $\eta^2 = 0.053$) and a significant two-way Valence Pair x Reach Direction interaction ($F_{(2,234)} = 15.04$, $p = 1.53e{-6}$, $\eta^2 = 0.040$). A simple main effect 1-factor RM-ANOVAs conducted to test for Valence Pair differences at each level of Reach

Direction revealed a significant main effect of Valence Pair for Leftward reaches ($F_{(2)} = 28.66$, *p* 7.34e-12), but not Rightward reaches. Subsequent pairwise comparisons between Leftward Valence Pair conditions showed a significant difference in standardized reaction time scores between all Valence Pair trial types different (High – High vs. Med – Med pairings: $p = 8.56e\text{-}4$; High – High vs. Low – Low pairings: $p = 5.42e\text{-}13$; Med – Med vs. Low – Low pairings: $p = 0.003$), with a linear trend in which High – High trials showed the fastest standardized reaction time scores, followed by Med – Med trials and finally Low – Low trials (see Figure 3.6 A, and Table A.3.4 for mean scores). These results suggest that our prediction of no difference in relative decision difficulty between matched pairs is supported when photos on the right were selected. However, leftward selections instead suggest that reaction time-expressed decision difficulty does indeed differ even when all photo pairs are matched in pleasantness, with selections between two photos rated lower in pleasantness showing the greatest difficulty. Ultimately, the expression of decision difficulty in reaction time appears to be dependent on the direction of movement (see Section 4.1 for discussion).

3.1.1.3.3.2 – Movement time analysis

A 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to the Photo Preference Matched-Pair standardized movement time scores revealed a main effect of Valence Pair ($F_{(1,234)} = 22.97$, $p = 2.16e\text{-}9$, $\eta^2 = 0.056$), a main effect of Reach Direction ($F_{(1,117)} = 16.52$, $p = 8.67e\text{-}5$, $\eta^2 = 0.026$), and a significant two-way Reach Direction x Device interaction ($F_{(2,117)} = 22.97$, $p = 2.16e\text{-}9$, $\eta^2 = 0.056$).

All pairwise comparison between levels of Valence Pair were significantly different (High – High vs. Med – Med pairings: p = 2.82e-4; High – High vs. Low – Low pairings: p = 2.65e-12; Med – Med vs. Low – Low pairings: p = 0.001), with means revealing a linear trend in which High – High trials showed the fastest standardized reaction time scores, followed by Med – Med trials and finally Low – Low trials with the slowest scores (see Figure 3.6 B; see Table A.3.4 for mean scores). These findings suggest that contrary to prior predictions, movement times also express differences in decision difficulty between matched photo pairs, with decision difficulty increasing as the pleasantness of the paired photos decreases.

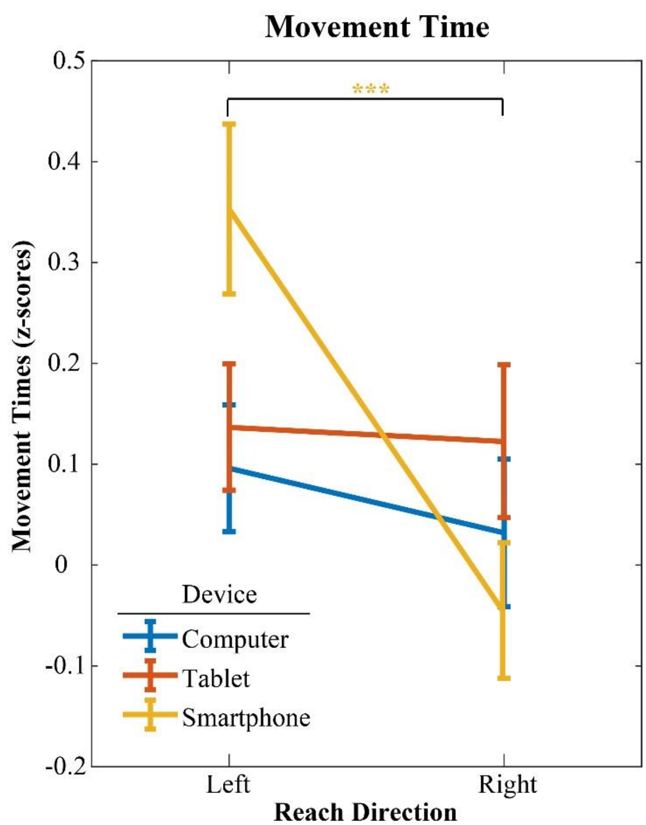

Figure 3.7 Photo Preference Matched-Pair mean standardized scores for Movement time measures, illustrating the interaction between Reach Direction and Device factors. Significant differences between Left and Right reach direction levels are indicated in the colour corresponding to the level of Device in which they were revealed. Error bars represent standard errors. Significance: * 0.01, ** 0.001, *** 0.0001.

To further understand the interaction between Reach Direction and Device, a simple main effect 1-factor RM-ANOVA was conducted to assess differences in Reach Direction at each level of Device. A main effect of Reach Direction was revealed at the level of Smartphone ($F_{(1)} = 27.97$, $p = 6.69e-6$), but not Computer or Tablet. At the level of Smartphone, trials with leftward reaches showed longer standardized movement time scores compared to rightward reaches (see Figure 3.7, and Table A.3.4 for mean scores). Device-dependent results are discussed further in Section 4.2.

3.1.1.3.3.3 – Trajectory analysis

A 3-factor Valence Pairing x Reach Direction x Device mixed-model ANOVA applied to the Photo Preference Matched-Pair standardized MAD scores revealed no main effects or interactions, suggesting no significant difference in trajectory curvature-expressed decision difficulty between trial conditions or devices used.

3.1.2 – Between-measures results

To assess the relationship between measures of decision difficulty, a within-participant correlation coefficient (r) was obtained for each combination of measures ($r_{MAD,MT}$ vs. $r_{MAD,RT}$ vs. $r_{MT,RT}$). It was predicted that each of these measures would be positively correlated, such that increases in one measure would be met by equivalent increases in the other. In line with this prediction, mean r values revealed trajectory curvature and movement time ($r_{MAD,MT}$) to be moderately positively correlated ($M_r = 0.33$, $SD = 0.25$). Contrary to our predictions, however, reaction time instead appeared to be weakly inversely correlated with both other measures, demonstrating small negative r values ($M_r = -0.083$, $SD = 0.15$ and $M_r = -0.039$, $SD = 0.20$ for $r_{MAD,RT}$ and $r_{MT,RT}$ correlations, respectively).

Participant average conditions correlations were then analyzed as a function of task and device differences to assess whether these relationships were consistent across task or device. We predicted that all measures would be positively correlated, and that this relationship would hold across task but not across device (which would reveal a diminished relationship between reaction time and both movement time and trajectory curvature, but not between the latter two).

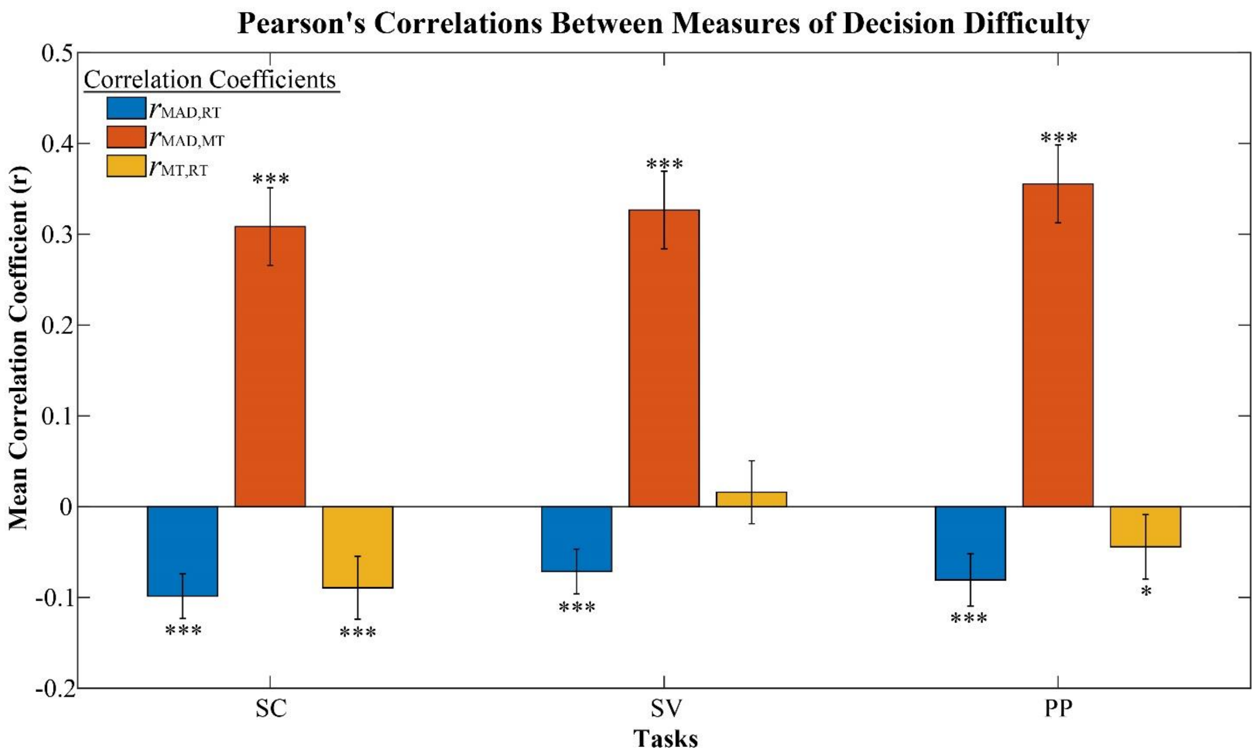

Figure 3.8 Mean Pearson's correlations (*r*) between measures of decision difficulty for Numeric-Size Congruency (SC), Sentence Verification (SV) and Photo Preference (PP) tasks. Degree of significance for mean correlations found to be significantly difference from 0 (no correlation) are indicated (* 0.01, ** 0.001, *** 0.0001). Error bars represent 95% confidence intervals.

The 3-factor Correlation Coefficient x Task x Device mixed model ANOVA testing for differences in the relationships between measures revealed a main effect of Correlation Coefficient ($F_{(2,234)}$ = 301.42, p = 1.55e-45, η^2 = 0.447) and a main effect of Task ($F_{(2,234)}$ = 7.70, p = 6.09, η^2 = 0.006). Figure 3.8 shows mean correlation coefficients for each level of

Correlation coefficient, separated by task. Pairwise comparison between levels of Correlation Coefficient showed a significant difference between $r_{\mathrm{MAD,MT}}$ correlation coefficients and both $r_{\mathrm{MAD,RT}}$ (p = 1.31e-59) and $r_{\mathrm{MT,RT}}$ (p = 7.43e-52) correlation coefficients. In both cases, $r_{\mathrm{MAD,MT}}$ was shown to be more positive than the other Correlation Coefficients. As with the reported mean correlations between measures, these findings contradict our prediction that all measures would be similarly positively correlated. Instead, positive $r_{\mathrm{MAD,MT}}$ correlations are distinct from $r_{\mathrm{MAD,RT}}$ and $r_{\mathrm{MT,RT}}$ correlations which are both negative and not significantly different from each other.

Pairwise comparisons between levels of Task revealed only a significant difference between Size Congruency and Sentence Verification (p = 5.09e-4), with Size Congruency revealing a smaller positive correlation when collapsed across all levels of Correlation Coefficient. No differences were revealed between Size Congruency and Photo Preference, or Sentence Verification and Photo Preference means. Despite these small differences in collapsed means, however, Task does not interact with the Correlation Coefficient factor, suggesting a consistent pattern within the correlated measures (that is, $r_{\mathrm{MAD,RT}}$ being positive and significantly different than $r_{\mathrm{MAD,RT}}$ and $r_{\mathrm{MT,RT}}$) independent of the task from which those measures were derived. Similarly, results showed no main effect of Device, nor did this factor interact with Correlation Coefficient suggesting this pattern to also be consistent across devices. That is, pre- and post-movement measures display an intricate relationship independent of their role in indexing task-specific decision difficulty – trajectory curvature and movement time, though positively correlated themselves, are inversely correlated with reaction times. This suggests a secondary pattern of results in which trajectory curvature and movement time decrease as reaction time increases (see Section 4.3 for a discussion of these results).

3.2 – Platform comparison

Results obtained from both the within-task and between-task analyses assessing differences as a function Platform widely replicated results previously reported and did not find any main effects or interaction with the between-subject factor of Platform. To avoid redundancies in result reporting, this data will not be discussed further.

4.0 – Discussion

The objective of this study was to assess whether metrics of decision difficulty - indexed through reaction time, movement time and trajectory curvature - remain consistent across decision domain, data collection device, and finally implementation platform. To accomplish this objective, we designed a three-task online experiment, each task replicating the design of prior mouse-tracked reach-decision experiments used to observe decision processes. These tasks included a Numeric-Size Congruency task (Faulkenberry *et al.,* 2016), a Sentence Verification task (Dale & Duran, 2011; Maldonado *et al.,* 2019) and a Photo Preference task (Koop & Johnson, 2013). Together these tasks spanned a range of decision domains from objective perceptual judgements to semi-subjective conceptual judgements and finally objective judgements of preference. Our experiment was then deployed across three devices (computers, tablets and smartphones) varying in size and interaction requirements using two testing platforms (Labvanced and Horizon) varying in their data export profiles and requiring customized data processing and cleaning strategies.

4.1 – Replication and extension

Broadly, task-specific results replicated previous mouse-tracked outcomes, with primary effects of interest reflected in a consistent pattern across all the measures of decision difficulty. Importantly, and most excitingly, all task-dependent decision difficulty effects were replicated independent of testing device or platform. This study therefore demonstrates the robustness of dynamic measures and offers seminal validation for the study of trajectory-tracked decision processes using small, portable devices. Further, these results suggest that trajectory-tracking techniques no longer need to be confined to a laboratory space, but that this data can be gathered online, from within people’s homes, and using any device that is readily available to them.

Adapted from earlier works by Faulkenberry and colleagues (2016), a Numeric-Size Congruency task presented participants with a numerical comparison between two digits varying in numerical value and physical size and required judgements to be made based solely on the dimension of numerical value. Classic size congruency effects, in which decision performance is impaired when the two dimensions differ (i.e., are incongruent in numeric and physical magnitude), are thought to reflect automatic processing of both characteristics despite only one being relevant to the task (and processing of irrelevant information may even be disadvantageous to optimal behaviour; Henik & Tzelgow, 1982). Through their use of a moused-tracked reach decision paradigm, Faulkenberry and colleagues (2016) extended prior reaction-time based evidence of this effect (Henik & Tzelgov, 1982; Santens, Gossen & Verguts, 2011) to movement time and mouse trajectory measures, supporting decision making models in which choice option processing and competition continues beyond reaction time rather than the decision processes being resolved prior to movement (Santens *et al.*, 2011; Song & Nakayama, 2009). The foundation of our predictions in the current study, choices made between incongruent digits were shown to incur longer reaction times (Henik & Tzelgov, 1982; Santens *et al.*, 2011; no differences in what the authors called "initiation times" were found by Faulkenberry *et al.*, 2016), longer movement times, and greater deviations towards the alternative choice in trajectory curvatures. Faulkenberry and colleagues further expanded on these effects by demonstrating that this size congruency effect interacts with numerical distance such that effects increase as numerical distance between digits increase and, as with reaction time (Schwarz & Ischebeck, 2003; Santens *et al.*, 2011), this modulation of effects is reflected in both movement time and mouse trajectories. The results of the current study provide further support for these effects, showing differences decision difficulty reflected in reaction time, movement time and mouse trajectory curvature between congruent and incongruent conditions in the Numeric-Size Congruency task (Figure 3.1). Specifically, as predicted, incongruent trials generated longer reaction times, longer movement times and trajectories with greater trajectory curvature.

Where our results differ from previous findings, however, is in the modulation of these effects by numerical distance. Our results show an interaction between the congruency effect and numerical value of the paired digits in which 8 – 9 pairings show the greatest congruency effects and differ most significantly from 1 – 2 and 2 – 8 pairings. This contradicts our prediction founded on Faulkenberry and colleagues (2016) results that 2 – 8 pairings would show the greatest effects with no difference between 1 – 2 and 8 – 9 pairings offering choices of the same numerical distance. These results may differ from those reported by Faulkenberry and colleagues because of our limiting our stimuli to a small subset of digits rather than an extensive range (1, 2, 8 and 9 in our case compared to 2, 3, 4, 5, 6, 7 and 8 in their second experiment examining numerical distance), washing out any relative effects. This does not, however, provide an explanation for a difference found between 1 – 2 and 8 – 9 pairings. Instead, two theories regarding perceived numerical processing and comparisons provide accounts compatible with our findings. The first stipulates that number comparison performance depends on the ratio of the two digits being compared, such that smaller number pairs are more quickly discriminated between because they present a larger ratio compared to larger number pairs with the same distance between them (Moyer & Landauer, 1967). The second theory suggests that numerically smaller digits are easier to process as they are more frequent (i.e., we are exposed to them more often in our daily lives; Dehaene and Mehler, 1992). While this theory was intended to explain a numerical distance effect (the further apart numbers are numerically the more processing advantage the smaller one has over the other), this can also be extended to explain performance differences between number pairs of the same numerical distance should the digits presented in one of the pairs be smaller (e.g., 1 – 2) compared to those in the other pair (8 – 9). Together, these theories provide a compelling argument for the difference observed between 1 – 2 and 8 – 9 pairings in the current study, wherein comparisons between the digits 8 and 9 reflect significantly greater decision difficulty.

A Sentence Verification task, replicating earlier works by Maldonado and colleagues (2019; adapted from Dale and Duran, 2011), required participants to verify statements varying in truth value and negation as true or false. Founded on the notion that negation acts as an operator on reading comprehension processes (Wason & Johnson-Laird, 1972), it is thought that its presence abruptly changes the predicted meaning of statements in the presence of insufficient context (e.g., simple, stand-alone statements; Dale & Duran, 2011). Reframed in terms of decision difficulty in the current study, mouse-tracked studies have shown negation to drive increases in trajectory curvature (Maldonado *et al.*, 2019; Dale & Duran, 2011). Maldonado and colleagues (2019) also demonstrated difference in trajectory curvatures between negated and non-negated statements to be greater when statements were true compared to when they were false. Trajectory curvature analyses in the current study replicate these interacting effects (see Figure 3.3). Additionally, these results were further exemplified in reaction time and movement time results, with the same conditions demonstrating greater trajectory curvature-expressed decision difficulty also showing longer reaction and movement times.

Finally, in a Photo Preference task adapted from Koop and Johnson (2013), participants were presented with pairs of photos varying in pleasantness (but matched in arousal) and asked to select the one they preferred. Seminally employed to validate the use of mouse-tracked measures to assess decision dynamics during purely subjective choice, preference was shown to parallel photo pleasantness ratings, producing trajectory curvatures reflecting greater decision difficulty as paired photo pleasantness increased in similarity. In our High-Chosen analysis (Section 3.1.1.3.1), we found a similar pattern of results within movement time and trajectory curvature measures of decision difficulty, with photos matched in pleasantness (High – High) showing greater decision difficulty than those not matched in pleasantness (High – Med and High – Low). However, despite a visibly linear trend of increased decision difficulty as photo similarity increases reflected in both movement time and trajectory curvature (see Figure 3.5), no significant difference was found between High – Med and High – Low pairs. This null effect may be due to an overvaluing of highly pleasant photos, such that relative pleasantness differences between choice alternatives (the difference between Med and Low photos in this case) are overlooked in the presence of a highly valued option (a High photo; Wispinski *et al.,* 2017). Additionally, despite a similar trend in the standardized means, no significant differences between photo pairs were revealed within reaction time measures, suggesting that in this particular case reaction times are less sensitive to changes in decision difficulty compared to movement times and trajectory curvature.

To assess whether differences in similarity were indeed the sole contributors to this effect, a second analysis examining differences in photo pairs matched at varying levels of pleasantness (Matched-Pair analysis, Section 3.1.1.3.2) was conducted. Contrary to prediction based on outcomes reported by Koop and Johnson (2013), movement time results revealed a linear trend in which movement time increases as the pleasantness of the matched pairs decreases (see Figure 3.6 B). Of note, however, is that these same results are not mirrored within reach trajectory measures, which show no differences between matched pairs. This dissociation between trajectory measures and movement time suggest that this effect might not arise as a reflection of decision difficulty (which would be reflected in both measures), but rather an outcome of increased vigor when moving towards something pleasant (as would be the case in High – High pairings) compared to something unpleasant (as would be the case in Low- Low pairings; Chapman, Gallivan, Wong, Wispinski & Enns, 2015). Reaction times showed a similar pattern of results to movement time, but only for leftward reaches (see Figure 3.3 A). These direction-dependent results could arise from preferential processing of stimuli presented on the right leading to decreased sensitivity to differences between rightward reaches, thus limiting the expression of pleasantness-driven decision difficulty differences to leftward reaches (Gallivan & Chapman, 2014).

4.2 – Device differences

Despite the evidence that indexes of decision difficulty are overwhelmingly consistent across devices, task outcomes also reflect device-dependent differences independent of these replication results. Here, we outline where those difference arise and offer rationale that accounts for these observed outcomes.

4.2.1 – Overview of device-contingent results

Within the Numeric-Size Congruency results, we see device differences reflected in factor interactions within each of the three measures of decision difficulty (see Section 3.1.1.1 for detailed results). Reaction time results reveal a modulation of reaction time-expressed decision difficulty differences between number pairs by device, such that differences between number pairs are more significant for computer-acquired data (Figure 3.2 A). Conversely, within movement time, we see an interaction between congruency-driven decision difficulty differences and device in which computer-acquired movement times show less significant differences between congruency conditions compared to tablet- and smartphone-acquire movement times

(Figure 3.2 B). Finally, trajectory curvature results reveal interaction between the side of numerically larger digit presentation (analogous with reach direction as only correct trials were included in the analysis) and device wherein right-sided presentations showed greater overall trajectory curvature compared to left-sided presentations, but this effect was not present during computer-based testing (Figure 3.2 C).

Device differences also emerge within each measure of decision difficulty in Sentence Verification task outcomes (see Section 3.1.1.2 for detailed results). First, reaction time results revealed an interaction between the truth value and polarity of statements and testing device used wherein greater truth value- and polarity-driven effects were expressed during computer use compared to tablet and smartphone use (Figure 3.4 A). Movement times reveal a modulation of polarity-driven differences by device, such that less significant effects were shown for computer-based testing (Figure 3.4 B). Finally, trajectory curvature results also show an interaction between truth value and device in which no truth-driven differences were uncovered for computer-based testing, but were for tablet- and smartphone-based testing (Figure 3.4 C).

Differences due to device use within the Photo Preference task arose only when photo pairs were matched in pleasantness (see Section 3.1.1.3.2 for detailed results) and appeared exclusively in movement time results. An interaction between device and reach direction was revealed, indicating a significant effect of reach direction within smartphone-based testing but not when computers or tablets were used (Figure 3.7).

4.2.2 – Speculated explanation for differences

To account for the outcomes described above, we first grouped the observed effects into three categories: those pertaining to directional bias differences, those pertaining to pre-movement processing differences (e.g., effects observed in reaction times), and those pertaining to post-movement processing differences (e.g., effects observed in movement times and trajectory curvatures).

To understand differences in directional biases observed through side of space factors (Presentation Side within the Numeric-Size Congruency task and Reach Direction within the Photo Preference task) and how they come to interact with device, we look towards an effect that consistently emerges in reach-decision studies: right-hand bias (Chapman & Gallivan, 2014). This direction-dependent bias arises from preferential processing of stimuli presented on the right, manifesting smaller condition-dependent differences in behavior (e.g., less variation in trajectory curvature) and generally faster movement times during rightward reaches. Notably, these effects are particularly prominent in real-world reaches towards three-dimensional targets (Chapman *et al.*, 2010b; Chapman & Gallivan, 2014). The presence of direction-dependent effects in trajectory curvature (Numeric-Size Congruency task) and movement time (Photo Preference) results obtained through use of touchscreen devices (smartphone and tablet or only smartphone, respectively; see Figure 3.2 C and 3.7) suggest that smartphones and tablets are better at replicating real movement effects. We therefore purport that directional effect differences between devices arise as a result of different user-interaction requirements enforcing different 'reach' biomechanics. Specifically, the swiping of a finger or sliding of a stylus across a screen surface appears to more closely resemble full-arm reaches during reach-decision tasks compared to mouse movements. Should a researcher be interested in more closely replicating real-world actions used to enact a choice, touchscreen-based testing may therefore provide a more biomechanically valid alternative to computer mouse tracking.

To understand devices differences manifesting in interactions with non-biomechanically driven factors (e.g., Congruency and Pairs in the Numeric-Size Congruency task and Truth and

Polarity in the Sentence Verification task), we first looked at effects observed in reaction times, thought to reflect decision processes occurring prior to movement. In both the Numeric-Size Congruency and Sentence verification task, we see an increase in the significance of condition differences within computer-acquired data compared to tablet- and smartphone-acquire data. In the Numeric-Size congruency task this manifested as a more significant Number Pairs main effect at the level of Computer compared to Tablet and Smartphone (Section 3.1.1.1.1, depicted in Figure 3.2 A) and in the Sentence Verification task this manifested as a more significant interaction between Truth and Polarity at the level of Computer compared to Tablet or Smartphone (Section 3.1.1.2.1, depicted in Figure 3.4 A). This pattern of results suggest that computer-based testing shows increased sensitivity to decision difficulty expressed within reaction time.

In examining device differences observed in post-movement measures (movement time and trajectory curvature), we see an inverse pattern emerge. Namely, we see a decrease in the significance of condition differences within computer-acquired data compared to tablet- and smartphone-acquire data. In the Numeric-Size Congruency task, this manifested as a less significant Congruency main effect at the level of Computer compared to Tablet and Smartphone (Sections 3.1.1.1.2 and 3.1.1.1.3, depicted in Figure 3.2 B). In the Sentence Verification task, a less significant effect was found at the level of Computer for both main effects of Polarity within movement time (Section 3.1.1.2.2, depicted in Figure 3.4 B) and Truth within trajectory curvature measures (Section 3.1.1.2.3, depicted in Figure 3.4 C). Together, these results support an increase in sensitivity to decision difficulty expressed within post-movement measures during tablet or smartphone use compared to computer use.

Grouped in this manner, the device-dependent outcomes observed in the current study evoke two primary forces driving device differences: biomechanics and measure-dependent sensitivity to decision difficulty. These forces manifest in across-device testing in the following ways:

1) Tablet- and smartphone-based movements reflect right-hand biases similar to real-world movements, whereas computer-based testing does not.
2) Computer-based testing shows more sensitivity to decision difficulty expressed in reaction times compared to tablet- and smartphone-based testing.
3) Tablet- and smartphone-based testing shows more sensitivity to decision difficulty expressed in movement times and trajectories compared to computer-based testing, displaying a distribution of sensitivity more akin to real-world reaching.

Importantly, however, these device-dependent effects do not impede successful capture of task-dependent decision difficulty effects (see Section 4.1).

Questions also remain as to whether differences in observed measure sensitivity to decision difficulty arise due to inherent device differences (e.g., size, vertical vs. horizontal distances, interaction methods) or if decision processes are fundamentally altered between devices (e.g., computer-based decisions induce greater seeping of decision processes into movement times, such that less of the decision is resolved during reaction time), or a combination of the two. For example, one could easily imagine that the greater horizontal distance between choice options presented on a landscape-oriented computer screen reduces the ease with which movements can be corrected if a change of mind occurs after reach onset compared to a smaller, portrait-oriented screen as there is a greater horizontal distance to travel back towards the alternative option and less vertical distance within which one can do so. The cost of mid-movement corrections could then produce a prioritization of decision resolution prior to movement, making reaction time a more sensitive measure of decision difficulty for computer-based reach-decisions compared to movement time or trajectory curvature. However, real-world mouse movements made to enact mouse cursor changes on a screen are physically very small. Perhaps then, these differences arise instead due to the nature of actual movements through space produced by the human body rather than the resultant movements expressed on a digital screen. Aligned with this argument, the smaller hand movement required to move a mouse cursor over large distances increases the opportunity for - and perhaps inevitability of - more ballistic responses. The smaller physical movements in space therefore require more decision to be resolved prior to movement initiation, as time during movement and real-world, physical space to express indecision is reduced. This particular tradeoff between reaction time and post-movement measures is reflected in the raw (non-standardized) measure means (Figure 4.1), which reveal reduced movement times during computer-use compared to tablet- or smartphone-use across each of the tasks. In line with this explanation, recall that movement time is thought to index both the complexity of the movement path (trajectory curvature, dependent on decision difficulty) and movement vigor (Dotan *et al.*, 2019) which was hypothesized to underlie a dissociation between trajectory curvature and movement times in the Photo Preference task (Section 4.1). Should differences in measure sensitivity during computer-based testing indeed arise from its reduced physical requirement during movement, this would again suggest that trajectory curvature and movement times are not analogous, only interconnected. Movement times may be more dependent on the ballistic nature of the movement than the difficulty of the decision and thus its sensitivity as a measure of decision difficulty is reduced with smaller, faster movements. In contrast, trajectory curvature may display reduced sensitivity to decision difficulty during computer use because more of the decision is resolved in reaction time prior to movement. As it is known that motor behaviors can inform cognitive processing (e.g., cognitive tuning; Strack, Martin & Stepper, 1988; Koop & Johnson, 2013), it is also reasonable to speculate that the requirement of a smaller movement

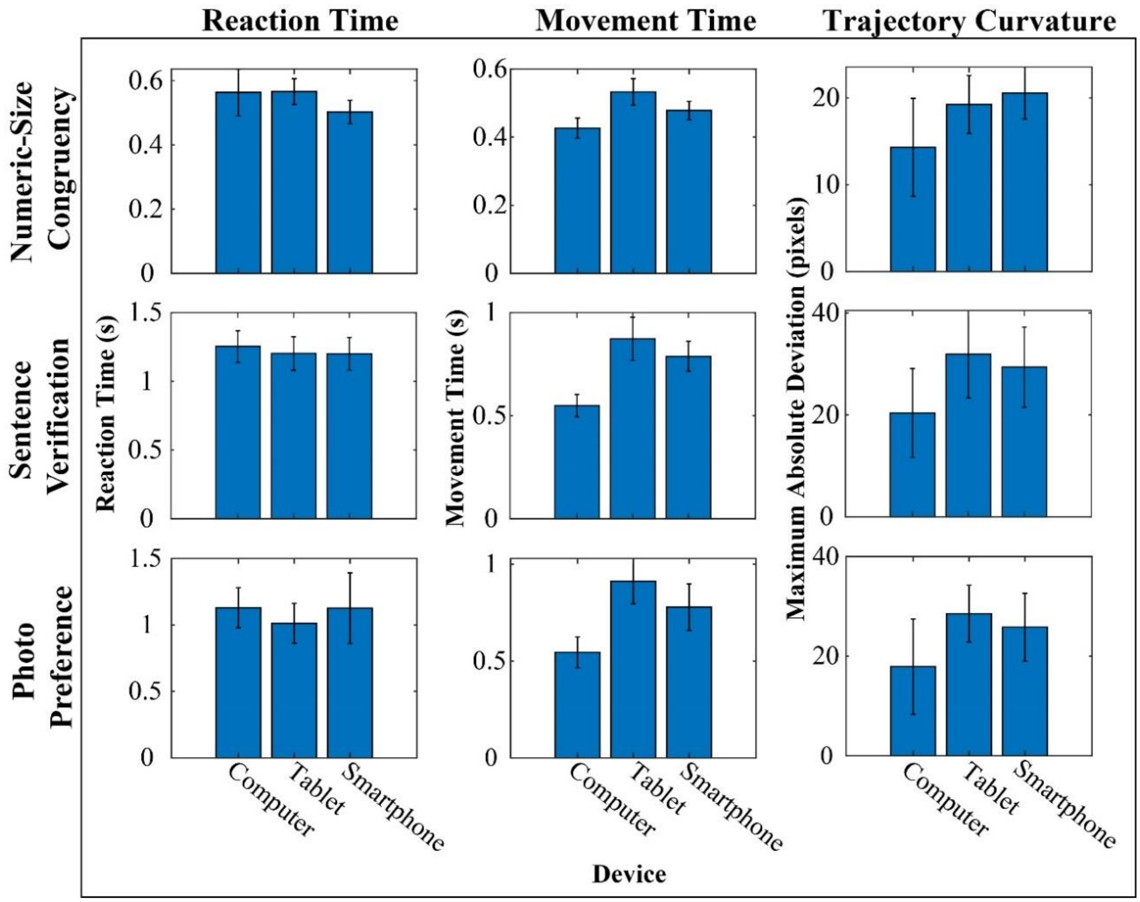

Figure 4.1 Raw (unstandardized) Reaction Time (left column), Movement Time (middle column) and Trajectory Curvature (right column) means for each task (Numeric-Size Congruency, top row; Sentence Verification, middle row; and Photo Preference, bottom row). Measure means are grouped by Device (Computer, Tablet and Smartphone). Error bars represent 95% confidence intervals.

would induce increased front-loading of decision processes prior to movement commencement (that is, a reciprocal relationship between shorter movement times and longer reaction times may exist).

4.3 – Within task differences between measures

While our replication of task results demonstrates an increase in reaction times, movement times, and trajectory curvatures as a function of increased decision difficulty, correlational analyses between these measures did not reveal a positive correlation between all measures as anticipated. Instead, while movement time and trajectory curvatures were positively correlated, both were shown to be inversely correlated with reaction time and this pattern was shown to be consistent between tasks and devices (Section 3.1.2).

An interesting dynamic between measures of decision difficulty is therefore revealed in which all three increase (i.e., show longer reaction times and movement times, or greater trajectory curvature) as decision difficulty increases, but separate from - or perhaps interacting with - these decision difficulty-driven changes, is an inverse relationship between reaction time and post-movement measures wherein increases in one mandate decreases in the other. Despite the apparent contradictions of these inter-dynamics, however, both are compatible within the rationale of an evidence accumulation model of decision making.

Within the models, evidence is noisily accumulated over time until a decision threshold has been reached (Wispinski *et al.,* 2020). More difficult decisions then arise from greater competition between choice which induce greater signal noise as evidence is accumulated for and against a choice option. Resolution of a difficult decision then requires more evidence to be accumulated before support for one option over another reaches the resolution threshold. This takes more time, and should these processes be ongoing after movement begins, is reflected in the physical pull experienced between options experienced during choice selection (Sullivan *et al.*, 2015; Stillman *et al*., 2020). Thus, as in the current study, more difficult decisions show greater reaction times, movement times, and trajectory curvatures. However, even if decision difficulty is unchanging (e.g., equal levels of evidence required for decision resolution), necessitating a specific amount of processing time, there is likely natural variation in reaction times. If decision processing requirements remain the same, but reaction time is reduced, decision processes must necessarily shift from reaction time into movement time. As the amount of processing required does not change, however, processing cannot seep into movement time without seeping out of reaction time, and vice-versa. This effect is what is reflected in the inversely correlation between reaction time and both movement time and trajectory curvature measures.

One can perhaps best appreciate these co-occurring effects by imaging decision making processes as being akin to a one-way train on a track of fixed length. The train begins on one end of the track (start of the trial), and on the other end is a gate (end of the movement). This gate is the point at which a decision is resolved. Between the start point and the gate, however, is also a county line (movement initiation, where reaction time meets movement time). The length of the track on either side of the county line determined by the demands of the decision task (e.g., movement initiation time constraints). The goal of the train is to amass enough coal (i.e., evidence in the context of decision making) to power its passage from one county into another, and through the gate. The degree of difficulty of the decision at hand dictates the number of train cars and the length of the train. The longer the train, however, the more coal is needed to power it. This necessarily takes more time - although this time can be reduced if the lumps of coal are larger or more quickly dispensed (i.e., the evidence provided is more salient or compelling). Coal can be amassed in either county and moving the train along the track does not reduce its length or the amount of coal required, only the amount amassed in the first county compared to the second. It is in this way that faster decisions can have more decision difficulty reflected in the movement - if less coal is collected in the first county, more coal must necessarily be collected in the second for the train to attain its goal. Importantly, this is true independent of the size of the train (i.e., the difficulty of the choice). The larger the train – or more difficult the decision – however, the more time is likely spent amassing coal in both counties. As such, while observing the proceedings of the train for the duration of its time moving along the track gives the most complete picture of the time course of events, indices of its length are accessible by observing the amount of coal amassed in either country independently. Similarly, while a complete account of the temporal dynamics of decision processes can only be found by observing the complete time course of a decision through reaction time and post-reaction time measures, each of those measures can independently provide indications of decision difficulty.

Overall, our analysis of both task-specific results and the relationship between measures of decision difficulty show that difficult decision have longer reaction times, longer movement times, and greater trajectory curvatures. However, for a given degree of decision difficulty (for example, all difficult decisions), if a specific decision displays a shorter reaction time, then residual evidence accumulation must take place during movement, giving rise to a longer movement time and greater reach curvature. Conversely, if a given decision has a longer reaction time, then more evidence can be accumulated before the reach starts, requiring a shorter movement time and showing a more direct reach.

4.4 – Implications and future directions

Our work significantly advances the emerging body of reach-decision research by successfully replicating task-dependent reach-trajectory effects using both testing platforms, and, less expectedly, across all three devices. Our accurate and informative online extension of classic mouse-tracking techniques to tablet and smartphone use cases shows the robustness of reach-decision paradigms and corroborates the sensitivity of dynamic indexes of decision difficulty (i.e., trajectory tracking). Further, given the ubiquitous use of smartphones, tablets and websites, our validation of these techniques as accessible outside the lab and impartial to testing device and platform differences breeds an impressively large market within research settings and industry alike, with many applications where detailed knowledge of decision dynamics could be useful (including domains such as corporate talent assessment and implicit bias).

Our work further indicates that where device differences do exist, they do not disrupt nor indicate pervasive decision difficulty effects but rather likely arise from biomechanical differences between device interactions and may reflect differences in the sensitivity of particular measures. For example, touchscreen devices may be more sensitive to decision difficulty expressed during movement while personal computers are more sensitive to decision difficulty expressed in reaction times. Should a researcher implement a task in which reaction time are highly controlled, it therefore follows that use of a touchscreen device may allow for the most sensitive capture of post-reaction time reach-paradigm measures. Conversely, if a task is known to produce highly condition-dependent reaction time effects, use of a computer-based testing system may allow for the most nuanced capture of those effects.

The present work also deepens our understanding of the relationship between popular measures of decision difficulty. Although each is thought to index decision processes at different time points over the course of a decision, our task-specific results were replicated within reaction times, movement times and reach trajectories, demonstrating them to each be powerful and accurate measures of decision difficulty. Our correlational analysis, however, further revealed a fixed relationship between the measures that was consistent between tasks and across platform and device: movement times and indices of trajectory curvature are positively correlated, increasing and decreasing in tandem, but both are inversely correlated with reaction time. These results align with current models of decision making in which decision processes beginning but not resolved prior to movement seep into movement, with ongoing choice competition reflected in movements made (and thus movement time; Wispinski *et al.*, 2020). In this way, depending on the choice at hand, decision processes can be flexibly adjusted between reaction time and movement time measures. Should more of the decision resolution occur in movement time, however, this indicates a shift of processing out of reaction time and manifests as an increase in movement times proportionate to the decrease in reaction times from what would normally be required to resolve the decision.

Finally, it is important to recognize some limitations of the current study and reaching-tracking paradigms general. While the current study brought trajectory tracking outside of the lab and onto devices participants likely make decisions using on a daily basis, it still employed a controlled design in which movement start and end points were largely regulated, and choices were limited to two options. Real-world decision-making and digital interactions are rarely so straight forward. Further research is needed to assess whether the consistency of outcomes found in the current study hold when decisions are made between more than two choices and are enacted using movement profiles more commonly seen in the real-world (e.g., scrolling down a screen to make a selection).

Additionally, while the current study explains device-driven result differences through differences in inherent measure sensitivity to decision difficulty, these claims are made on limited evidence accrued between tasks. Although a detailed discussion and analysis of the predictive ability of the measures collected are beyond the scope of the current study, we believe predictive modelling of the reported measures of decision difficulty as a function of device might produce insights into whether the sensitivity of particular measures are indeed dependent on testing device (for example, whether reaction time more is sensitive to - and thus more predictive of – differences in decision difficulty during computerized testing compared to tablet or smartphone-based testing).

As eye-tracking has evolved such that capture of reliable data is now possible via webcam, and these technologies being extended to the cameras of portable devices (e.g., smartphone camera), future directions that may also yield important results regarding decision making processes and the consistency of metrics of decision difficulty between tasks and devices include measuring eye-movements.

4.5 – Conclusion

Computer mouse cursor movements reflect underlying cognitive processes, and, when measured during a computerized choice between options, reaction time, movement time and mouse trajectory curvature can serve as indexes of decision difficulty and provide insight into the evolution of decisions over time (Song & Nakayama, 2009; Stillman *et al.,* 2020). Employing three previously studied reach-decision tasks, the current study aimed to assess whether within-participant metrics of decision difficulty remain consistent across decision domains varying in choice stimuli, objectivity and processing requirement, data collection devices varying size and user-interaction requirements (e.g., mouse-based interactions to touchscreen use) and implementation platforms requiring individualized data processing and cleaning strategies. Study outcomes add to the current body of literature that investigates decision processes using mouse-tracking techniques by providing the first replication of classically computer-acquired task-specific results using two platforms, and, most excitingly, two portable, touchscreen devices: tablets and smartphones. Where differences between data collection device did arise, they were independent of primary task effects and served to suggest that tablet- and smartphone-based testing shows more sensitivity to decision difficulty expressed in post-reaction time measures, displaying directional biases and sensitivity distributions more reminiscent of real-world reaches compared to computer-based testing. Contrasting its touchscreen-based counterparts, computer-based testing appeared to show more sensitivity to decision difficulty expressed in reaction time, suggesting a flexible distribution of decision processes between pre- and post-reaction time measures that is dependent on the device in use and the movements required to enact a choice using it. Lending further support to the flexible and continuous nature of decision processes, between-measure correlations revealed a decision-dependent effect existing in parallel to task-dependent choice competition effects: while increased decision difficulty is reflected in increased reaction times, movement times and trajectory curvature, an inverse relationship also exists between these measures such that, for a given degree of decision difficulty and in the presence of natural variations in processing dynamics, movement time and trajectory curvature increase as reaction time decreases. Together, these results provide support for models of decision making in which decision processes continue to unfold after movements to enact a choice have been initiated, and further suggest that these processes are flexibly adjusted along the time course of a decision both when choice execution requirements change (e.g., device interaction methods change) and when decision domain and difficulty remain consistent. Importantly, however, despite this flexibility, these interacting measures remain powerful indicators of decision difficulty and impervious to changes in testing platform or data collection device. This study therefore serves to reinforce our understanding of the intimate connection between cognition and movement, and further emphasizes the power of trajectory tracking methods in unveiling the dynamism of decision processes.

References

Arnold, B.F., Hogan, D.R., Colford, J.M., & Hubbard, A.E. (2011) Simulation methods to estimate design power: an overview for applied research. *BMC Medical Research Methodology*, *11*, 94. **https://doi.org/10.1186/1471-2288-11-94**

Baumann, M.A., Fluet, M.C., & Scherberger, H. (2009). Context-specific grasp movement representation in the macaque anterior intraparietal area. *Journal of Neuroscience, 29*(20), 6436–6448. **https://doi.org/10.1523/JNEUROSCI.5479-08.2009**

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22*(1), 23-36. **https://doi.org/10.2307/1909829**

Chapman, C. S., Gallivan, J. P., Wong, J. D., Wispinski, N. J., & Enns, J. T. (2015). The snooze of lose: Rapid reaching reveals that losses are processed more slowly than gains. *Journal of Experimental Psychology: General, 144*(4), 844–863. **https://doi.org/10.1037/xge0000085**

Chapman, C.S., & Goodale, M.A. (2010). Obstacle avoidance during online corrections. *Journal of Vision, 10*(11), 1-14. **https://doi.org/10.1167/10.11.17**

Chapman, C.S., Gallivan, J.P., Wood, D.W., Milne, J.L., Culham, J.C., & Goodale, M.A. (2010a). Short-term motor plasticity revealed in a visuomotor decision-making task. *Behavioral Brain Research, 214*(1), 130- 134.

Chapman, C.S., Gallivan, J.P., Wood, D.W., Milne, J.L., Culham, J.C., & Goodale, M.A. (2010b). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition, 116*(2), 168-176.

Chen, X. & Stuphorn, V. (2015). Sequential selection of economic good and action in medial frontal cortex of macaques during value-based decisions. *eLife, 4,* e09418. **https://doi.org/10.7554/eLife.09418**

Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinions in Neurobiology, 22*(6), 927-936. **https://doi.org/10.1016/j.conb.2012.05.007**

Cisek, P., & Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron, 45*(5), 801-814. **https://doi.org/10.1016/j.neuron.2005.01.027**

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33*, 269-298. **https://doi.org/10.1146/annurev.neuro.051508.135409**

Cramer, A. O., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P., Waldorp, L. J., & Wagenmakers, E. J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647. **https://doi.org/10.3758/s13423-015-0913-5**

Cunningham, W.A., Dunfield, K.A., & Stillman, P.E. (2013). Emotional states from affective dynamics. Emotion Review, *5*, 344-355.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science, 35*(5), 983-996. **https://doi.org/10.1111/j.1551-6709.2010.01164.x**

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43*(1), 1–29. **https://doi.org/10.1016/0010-0277(92)90030-l**

Dotan, D., Pinheiro-Chagas, P., Roumi, F. A., & Dehaene, S. (2019). Track it to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, *23*(12), 1058-1070. **https://doi.org/10.1016/j.tics.2019.10.002**

Erb, C.D., Moher, J., Song, J.H., & Sobel, D.M. (2018). Numerical cognition in action: reaching behavior reveals numerical distance effects in 5- to 6-year-olds. *Journal of Numerical Cognition*, *4*(2), 286-296. **https://doi.org/10.5964/jnc.v4i2.122**

Farmer, T.A., Cargill, S.A., Hindy, N.C., Dale, R., & Spivey, M.J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science, 31*(5), 889-909. **https://doi.org/10.1080/03640210701530797**

Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Canadian Journal of Experimental Psychology, 68*(3), 147–151. **https://doi.org/10.1037/cep0000021**

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica, 163,* 114–123. **https://doi.org/10.1016/j.actpsy.2015.11.010**

Feiveson, A.H. (2002). Power by simulation. *Stata Journal, 2,* 107-124. **https://doi.org/10.1177/1536867X0200200201**

Freeman J. B., Ambady N., Midgley K. J., Holcomb P. J. (2011). The real-time link between person perception and action: brain potential evidence for dynamic continuity. *Society for Neuroscience*, *6*, 139–155. **https://doi.org/10.1080/17470919.2010.490674**

Freeman, J. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226-41. **https://doi.org/10.3758/BRM.42.1.226**

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science, 27*(5), 315–323. **https://doi.org/10.1177/0963721417746793**

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: split-second social perception. *Trends in Cognitive Sciences, 20*(5), 362-374. **https://doi.org/10.1016/j.tics.2016.03.003**

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in psychology*, *2*, 59. **https://doi.org/10.3389/fpsyg.2011.00059**

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science, 27*(4), 502-517. **https://doi.org/10.1177/0956797615627418**

Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in Neuroscience*, *8*, 215. **https://doi.org/10.3389/fnins.2014.00215**

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience, 19*(9), 519–534. **https://doi.org/10.1038/s41583-018-0045-9**

Gallivan, J.P., Chapman, C.S., Wood, D.W., Milne, J.L., Ansari, D.A., Culham, J.C., & Goodale, M.A. (2011) One to four, and nothing more: Non-conscious parallel object individuation in action. *Psychological Science, 22* (6). **https://doi.org/10.1177/0956797611408733**

Ghez, C., Favilla, M., Ghilardi, M.F., Gordon, J., Bermejo, R., and Pullman, S. (1997). Discrete and continuous planning of hand movements and isometric force trajectories. *Experimental Brain Research*. 115, 217–233.

Ghez, C., Gordon, J., Ghilardi, M. F., Christakos, C. N., & Cooper, S. E. (1990). Roles of proprioceptive input in the programming of arm trajectories. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 837–847. **https://doi.org/10.1101/sqb.1990.055.01.079**

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations, 18*(3), 384–401. **https://doi.org/10.1177/1368430214538325**

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory and Cognition,* 10, 389–395. **https://doi.org/10.3758/BF03202431**

Huk, A., & Shadlen, M. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of Neuroscience, 25*, 10420-10436.

Hunt, L. T., Malalasekera, W., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature neuroscience, 21*(10), 1471–1481. **https://doi.org/10.1038/s41593-018-0239-5**

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. **https://doi.org/10.1037/0033-295X.87.4.329**

Khaw, M. W., Glimcher, P. W., & Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences of the United States of America, 114*(48), 12696–12701. **https://doi.org/10.1073/pnas.1715293114**

Kieslich, P., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2019). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods.* 52. **https://doi.org/10.3758/s13428-019-01228-y**

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, 67, 151–185. **https://doi.org/10.1016/j.cogpsych.2013.09.001**

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857. **https://doi.org/10.1073/pnas.1101328108**

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8.* University of Florida, Gainesville, FL.

Link, S., & Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*, 77-105.

Maldonado, M., Dunbar, E., Chemla, E. (2019). Mouse tracking as a window into decision making. Behavior Research Methods. *Psychonomic Society, Inc. 51*(3), 1085-1101.

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science (New York, N.Y.)*, *211*(4477), 77–80. **https://doi.org/10.1126/science.7444452**

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action Dynamics Reveal Parallel Competition in Decision Making. *Psychological Science*, *19*(1), 22–24. **https://doi.org/10.1111/j.1467-9280.2008.02041.x**

Moher, J., & Song, J.H. (2019). A comparison of simple movement behaviors across three different devices. *Attention, Perception and Psychophysics*. 81, 2558-2569. **https://doi.org/10.3758/s13414-019-01856-8**

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. **https://doi.org/10.1038/2151519a0**

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*(7090), 223–226. **https://doi.org/10.1038/nature04676**

Palmer, J., Huk, A.C., & Shadlen, M.N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 376-404. **https://doi.org/10.1167/5.5.1**

Pinheiro-Chagas, P., Dotan, D., Piazza, M., & Dehaene, S. (2017). Finger Tracking Reveals the Covert Stages of Mental Arithmetic. *Open mind: discoveries in cognitive science, 1*(1), 30–41. **https://doi.org/10.1162/OPMI_a_00003**

Platt, M., & Glimcher, P. (1999). Neural correlates of decision variables in parietal cortex. *Nature. 400,* 233–238. **https://doi.org/10.1038/22268**

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current opinion in neurobiology, 20*(2), 262–270. **https://doi.org/10.1016/j.conb.2010.03.001**

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922. **https://doi.org/10.1162/neco.2008.12-06-420**

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. **https://doi.org/10.1111/1467-9280.00067**

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review,* 9, 438–481. **https://doi.org/10.3758/BF03196302**

Rosenbaum, D. (2005). The Cinderella of psychology: The neglect of motor control in the science of mental life and behavior. *The American psychologist*, *60*(4). 308-17. **https://doi.org/10.1037/0003-066X.60.4.308**

Santens, S., Goossens, S., & Verguts, T. (2011). Distance in motion: Response trajectories reveal the dynamics of number comparison. *PLoS One, 6*(9), e25429. **https://doi.org/10.1371/journal.pone.0025429**

Schoemann, M., Lüken, M., Grage, T., Kieslich, P., & Scherbaum, S. (2019). Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods, 51,* 2356–2377. **https://doi.org/10.3758/s13428-018-1179-4**

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2020). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*. Advance online publication. **https://doi.org/10.3758/s13423-020-01851-3**

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science, 26*(5), 442–450. **https://doi.org/10.1177/0963721417708229**

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance, 29*(3), 507–522. **https://doi.org/10.1037/0096-1523.29.3.507**

Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron, 90*(5), 927–939. **https://doi.org/10.1016/j.neuron.2016.04.036**

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*(2), 135–168. **https://doi.org/10.1016/0022-2496(88)90043-0**

Sobel, K. V., & Puri, A. M. (2018). Size congruity influences visual search via the target template. *Acta Psychologica,* 183, 66-74. **https://doi.org/10.1016/j.actpsy.2018.01.003**

Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in cognitive sciences, 13*(8), 360–366. **https://doi.org/10.1016/j.tics.2009.04.009**

Spivey, M. (2008). The Continuity of Mind. 10.1093/acprof:oso/9780195170788.001.0001.

Spivey, M.J., Grosjean, & M., Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29) 10393-10398. **https://doi.org/10.1073/pnas.0503903102**

Stillman, E., Krajbich, I. & Ferguson, M. (2020). Using dynamic monitoring of choices to predict and understand risk preferences. *Proceedings of the National Academy of Sciences, 117*(50) 31738-31747. **https://doi.org/10.1073/pnas.2010056117**

Stillman, P.E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Science, 22,* 531–543. **https://doi.org/10.1016/j.tics.2018.03.012**

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology, 54*(5), 768–777. **https://doi.org/10.1037//0022-3514.54.5.768**

Szászi, B., Palfi, B., Szollosi, A., Kieslich, P., & Aczel, B. (2018). Thinking dynamics and individual differences: Mouse-tracking analysis of the denominator neglect task. *Judgment and Decision Making*, *13*(1), 23-32.

Wason, P. C. (1959). The processing of positive and negative information. *The Quarterly Journal of Experimental Psychology, 11*(2), 92–107. **https://doi.org/10.1080/17470215908416296**

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: bridging the gap between decision making and action. *Annals of the New York Academy of Sciences, 1464*(1), 30–51. **https://doi.org/10.1111/nyas.13973**

Wispinski, N. J., Truong, G., Handy, T. C., & Chapman, C. S. (2017). Reaching reveals that best-versus-rest processing contributes to biased decision making. *Acta Psychologica, 176*, 32–38. **https://doi.org/10.1016/j.actpsy.2017.03.006**

Yarbus, A. L. (1967). Eye Movements and Vision. New York: Plenum Press. **https://doi.org/10.1007/978-1-4899-5379-7**

Zgonnikov, A. Aleni, A., Piiroinen, P. T., O'Hora, D., & di Bernardo M. (2017). Decision landscapes: visualizing mouse-tracking data. *Royal Society Open Science, 4*(11). **http://doi.org/10.1098/rsos.170482**

Appendix

A.1 - Participant survey responses

Table A.1.1 Table of demographic survey responses

		Disclosed Responses (count)			
		Labvanced			Horizon
Survey Questions	**Response Options**	**Computer**	**Tablet**	**Smartphone**	**Computer**
		N = 83	*N = 78*	*N = 78*	*N = 40*
What gender do you identify as?	Male	56	50	52	14
	Female	25	27	25	24
	Prefer not to say	2		1	
	Other *(custom response)*		1		2
What is your age?	*(custom response)*	range [21-65]	range [23-58]	range [24-57]	range [17-43]
What hand do you typically write with (i.e. your dominant hand)?	Right	76	66	71	35
	Left	6	12	7	5
Do you have normal or corrected to normal vision? (vision correction may be via surgery, glasses, contacts, etc.)	Yes	79	76	73	36
	No	4	2	5	4
Is English your first language?	Yes	80	76	76	22
	No	2	2	2	18
If English is not your first language, how old were you when you learned English?	*(custom response)*	4,23	3,4	3.5	3,4,5,6,8,9,20
How would you rate your English reading comprehension skills? (1-5)	1			1	
	2				
	3				
	4	2	2	3	7
	5	81	76	74	33
How many hours a week do you play video games?	0	2	2	2	1
	Less than 1	6	6	3	13
	1 - 3	16	13	15	6
	4 - 6	16	14	14	6
	7 - 10	12	17	14	1
	11 - 15	7	6	10	2
	16 - 20	11	5	8	
	20+	13	14	12	1
How many hours a week do you participate in activities requiring coordinated hand - eye movement? (e.g. instrument playing, catching and throwing, swinging, dribbling, etc.)	0	8	1	4	4
	Less than 1	21	14	10	7
	1 - 3	18	26	24	17
	4 - 6	12	14	11	8
	7 - 10	4	7	11	3
	11 - 15	6	3	4	1
	16 - 20	2	5	3	
	20+	12	7	11	

Table A.1.2 Table of device survey responses

Survey Questions	Response Options	Disclosed Responses (count) Labvanced	Horizon
Computer		*N = 83*	*N = 40*
What device are you using to complete this experiment?	Desktop	69	37
	Laptop	14	3
Desktop Computer			
What brand of desktop computer are you using?	Apple	1	5
	Acer	1	5
	Asus	9	1
	Dell	9	6
	HP	13	4
	Lenovo	1	3
	Microsoft	5	2
	I don’t know	3	1
	Other *(custom response)*	25*	9**
What brand is your monitor?	Apple	3	4
	Acer	11	5
	Asus	10	3
	Dell	13	8
	HP	7	5
	Lenovo	2	
	Microsoft		
	I don’t know	2	1
	Other *(custom response)*	21***	11****
What size is your monitor?	22 inches	10	5
	24 inches	28	13
	27 inches	19	6
	34 inches	4	
	I don’t know	2	10
	Other *(custom response)*	6*****	
What operating system does your computer use?	Windows	64	30
	Mac OS	2	5
	Linux	1	
	Chrome OS	1	2
	I don’t know	1	
	Other *(custom response)*		

Note: table continued on next page

Table A.1.2 *(continued)*

How are you choosing to interact with your computer?	Wired mouse	52	29
	Wireless mouse	17	8
	Trackpad		
	Touchscreen		
	Other *(custom response)*		
Laptop Computer			
What brand of laptop are you using?	Apple		
	Acer	4	
	Asus		
	Dell	4	1
	HP	6	1
	Lenovo		
	Microsoft		
	I don't know		
	Other *(custom response)*		1†
What model/series of laptop from the brand are you using? (e.g., MacBook Air, Asus Vi-voBook, Microsoft SurfaceBook, etc.) Please be as specific as posible.	*(custom response)*	††	†††
What is your laptop's screen size? (round to nearest option if necessary)	11 inches	1	1
	14 inches	2	1
	15 inches	5	
	16 inches	1	1
	17 inches	5	
	I don't know		
	Other *(custom response)*		
What operating system does your laptop use?	Windows	13	3
	Mac OS		
	Linux		
	Chrome OS	1	
	I don't know		
	Other *(custom response)*		
How are you choosing to interact with your laptop?	Wired mouse	4	1
	Wireless mouse	10	2
	Laptop touchpad		
	Wired/wireless trackpad		
	Touchscreen		
	Other *(custom response)*		

Note: table continued on next page

Table A.1.2 *(continued)*

Tablet		*N = 78*
What brand of tablet are you using?	Samsung	38
	Apple	
	Google	6
	Amazon	25
	Microsoft	
	HTC	
	Huawei	
	I don't know	2
	Other *(custom response)*	7‡
What is the make and model/series of the tablet you are using? (e.g., 10.2 Apple iPad Pro, Samsung Galaxy Tab S6, Google Pixel Slate, etc.). Please be as specific as possible.	*(custom response)*	‡‡
What is your tablet's screen size? (round to the nearest option if necessary)	7 inches	15
	8 inches	22
	9 inches	4
	10 inches	28
	11 inches	4
	I don't know	4
	Other *(custom response)*	1‡‡‡
What operating system does your tablet use?	Android	76
	Apple iOS	
	Harmony OS	
	Windows	
	I don't know	2
	Other *(custom response)*	
How are you choosing to ineract with your tablet?	Wired mouse	
	Wireless mouse	
	Finger/thumb on touchscreen	73
	Stylus on touchscreen	5
	Trackpad	
	Other *(custom response)*	
Smartphone		*N = 78*
What brand of smartphone are you using?	Samsung	45
	Apple	
	Google	4
	LG	10
	Nokia	2
	Motorola	6

Note: table continued on next page

Table A.1.2 *(continued)*

	Huawei	2
	I don't know	
	Other *(custom response)*	9§
What is the make and model/series of the smartphone you are using? (e.g., Samsung Galaxy S20+, Apple iPhone 11 Pro, Huawei P30 Lite, etc.). Please be as specific as possible.	***(custom response)***	§§
What operating system does your smartphone use?	Android	77
	Apple iOS	
	Harmony OS	
	Bada	
	I don't know	1
	Other *(custom response)*	
How are you choosing to interact with your smartphone?	Finger on touchscreen	43
	Thumb on touchscreen	34
	Stylus on touchscreen	1
	Other *(custom response)*	

Disclosed custom responses (note, only unique responses are reported here):

* Custom build, CyberPowerPC, MAINGEAR, iBUYPOWER

** Custom build, CORSAIR, CyberPower

*** AOC, BenQ, LG, Panasonic (television), Samsung, ViewSonic, Planar

**** BenQ, LG, Samsung

***** 16 inches, 20 inches, 30 inches, 32 inches, 42 inches

† RCA

†† Acer Chromebook 15, Acer Nitro 5, Acer Aspire 3, Dell Inpsiron 7300, HP ZBook, Dell Inpsiron 15 5570, Dell Inspiron 15 5000, Dell XPS 15, HP EliteBook, HP Pavilon

††† RCA Cambio Windows 2-in-1 Tablet/Laptop, Dell Vostro 3550, HP Spectre x360

‡ Acer, Lenovo, RCA, TECNO

‡‡ Acer Iconia One 7, Amazon Fire, Amazon Fire HD, Amazon Fire 7, amazon Fire 8, Amazon Fire 8 HD, Amazon Fire 10, Amazon Fire 10 HD, Amazon Kindle, Amazon Kindle Fire, Amazon Kindle Fire 5, amazon Kindle Fire HD, Fusion, Google Nexus 7, Google Pixel Slate, Lenovo M10 Plus, Lenovo Tab 4, ONN Tablet, RCA Galileo, Samsung Galaxy Tab 2, Samsing Galaxy Tab 4, Samsung Galaxy Tab A, Samsung Galaxy Tab A 10.1, Samsung Galaxy Tab A6, Samsung Galaxy Tab A7, Samsung Galaxy Tab A8, Samsung Galaxy Tab E, Samsung Galaxy Tab S10, Samsung Galaxy Tab S2, Samsung Galaxy Tab S3, Samsung Galaxy Tab S6, Samsung Galaxy Tab S6 Lite, Samsung Galaxy Tab S7, TECNO DroiPad, I don't know

‡‡‡ 12.4

§ Asus, BLU, Realme, TECNO, OnePlus, Xiaomi

§§ Asus ROG Phone 2, Blu G90 Pro, Google Pixel 1, Google Pixel 3, Google Pixel 3a XL, Huawei P20, LG G3, LG G6, LG K20, LG Revolution, LG Stylo, LG Stylo 4, LG Stylo 5, LG, Tribute HD, Motorola Moto G Power, Motorola Moto G4, Motorola Moto G7 Play, Motorola Moto G7 Power, Motorola Moto X4, Nokia 5.3, Nokia 6.1, OnePlus 6, OnePlus 8T, Realme Note 6 Pro, Samsung Galaxy, Samsung Galaxy A10, Samsung Galaxy A10E, Samsung Galaxy A20, Samsung Galaxy A71, Samsung Galaxy J3, Samsung Galaxy J3V, Samsung Galaxy J7, Samsung Galaxy Note 10+, Samsung Galaxy Note 20 Ultra, Samsung Galaxy S10, Samsung Galaxy S10+, Samsung Galaxy S10E, Samsung Galaxy S20, Samsung Galaxy S20 FE, Samsung Galaxy S20 Ultra, Samsung Galaxy S6, Samsung Galaxy S7, Samsung Galaxy S7E, Samsung Galaxy S8, Samsung Galaxy S8+, Samsung Galaxy S9, Samsung Galaxy S9+, Samsung Galaxy S20+, Techno Camon 15 CD7, Xiaomi Redmi Note 5, Xiaomi Redmi Note 8 Pro, Xiaomi Redmi Note 9

Table A.1.3 Table of device use survey responses

Survey Questions	ResponseOptions	Disclosed Responses (count)			
		Labvanced			Horizon
		Computer	Tablet	Smartphone	Computer
		N = 83	*N = 78*	*N = 78*	*N = 40*
Which hand are you using to make responses on your device?	Left	12	8	8	5
	Right	71	70	70	35
How many hours a week do you spending using this device?	0 - 5		20	7	16
	6 - 10	4	20	15	3
	11 - 15	3	12	14	3
	16 - 20	4	12	9	2
	21 - 25	9	8	9	2
	26 - 30	9	1	7	3
	31 - 35	7	1	4	6
	36 - 40	4	1	5	
	40+	43	2	8	5
How many hours a week do you typically use this device for the following activities:					
Surfing the internet	0		3	1	10
	Less than 1	6	16	11	7
	1 - 3	17	30	17	9
	4 - 6	9	12	18	4
	7 - 10	19	7	11	2
	11 - 15	11	3	5	2
	16 - 20	5	2	12	2
	20+	16	4	3	3
Playing games	0	6	10	15	23
	Less than 1	7	24	21	7
	1 - 3	19	22	21	4
	4 - 6	18	6	13	3
	7 - 10	8	6	2	1
	11 - 15	8	6	4	
	16 - 20	7	3	1	
	20+	10		1	2
School (e.g., research, assignments, writing, etc.)	0	68	62	64	14
	Less than 1	3	11	5	4
	1 - 3	5	2	4	6
	4 - 6			2	4
	7 - 10	3		1	
	11 - 15	2		1	4
	16 - 20				1
	20+	1		1	2

Note: table continued on next page

Table A.1.3 *(continued)*

Work (e.g. research, teaching, word processing, design, etc.)	0	8	46	21	14
	Less than 1	6	10	16	5
	1 - 3	7	8	22	5
	4 - 6	6	3	8	5
	7 - 10	10	4	3	3
	11 - 15	5	2	4	4
	16 - 20	7	2	3	2
	20+	34	1	1	2
Personal communication (e.g., talking with friends and family, etc.)	0	8	29		15
	Less than 1	18	23	12	5
	1 - 3	27	10	30	15
	4 - 6	12	7	12	2
	7 - 10	7	7	8	2
	11 - 15	6	1	8	
	16 - 20	1		4	
	20+	4		4	
School/work communication (e.g., email, phone calls, video conferencing, etc.)	0	16	52	26	9
	Less than 1	26	15	17	14
	1 - 3	17	8	20	7
	4 - 6	10		8	5
	7 - 10	4	2	1	2
	11 - 15	3		1	2
	16 - 20	3		4	
	20+	4		1	

A.2 - Sentence Verification statement stimuli

Table A.2.1 Table of Sentence Verification statement stimuli

True		False	
Positive	**Negative**	**Positive**	**Negative**
Elephants are large.	Elephants are not small.	Elephants are small.	Elephants are not large.
Cars have tires.	Cars do not have wings.	Cars have wings.	Cars do not have tires.
Grass is green.	Grass is not blue.	Grass is blue.	Grass is not green.
Ice is cold.	Ice is not warm.	Ice is warm.	Ice is not cold.
Boulders are heavy.	Boulders are not light.	Boulders are light.	Boulders are not heavy.
Rocks are hard.	Rocks are not soft.	Rocks are soft.	Rocks are not hard.
Dogs bark.	Dogs do not meow.	Dogs meow.	Dogs do not bark.
Apples are fruit.	Apples are not vegetables.	Apples are vegetables.	Apples are not fruits.
Candy is sweet.	Candy is not salty.	Candy is salty.	Candy is not sweet.
Knives are sharp.	Knives are not blunt.	Knives are blunt.	Knives are not sharp.
Villains are evil.	Villains are not kind.	Villains are kind.	Villains are not evil.
Fire is hot.	Fire is not cold.	Fire is cold.	Fire is not hot.
The sun is bright.	The sun is not dim.	The sun is dim.	The sun is not bright.
Rockets are fast.	Rockets are not slow.	Rockets are slow.	Rockets are not fast.
Car horns are loud.	Car horns are not quiet.	Car horns are quiet.	Car horns are not loud.
Turtles are slow.	Turtles are not fast.	Turtles are fast.	Turtles are not slow.
Giraffes are tall.	Giraffes are not short.	Giraffes are short.	Giraffes are not tall.
Garbage smells bad.	Garbage does not smell good.	Garbage smells good.	Garbage does not smell bad.
Heroes are helpful.	Heroes are not useless.	Heroes are useless.	Heroes are not helpful.
Kids like to play.	Kids do not like to do homework.	Kids like to do homework.	Kids do not like to play.
Diamonds are shiny.	Diamonds are not dull.	Diamonds are dull.	Diamonds are not shiny.

A.3 - Tables of task-specific effects and means

Table A.3.1 Table of Numeric-Size Congruency means

Reaction Time (z-scored)

Interaction** Pairs x Congruency

		Congruency				
		Congruent		Incongruent		F
Pairs	1v2	-0.199	0.381	0.028	0.395	***
		Congruent <<< Incongruent				
	2v8	-0.101	0.379	0.009	0.366	***
		Congruent << Incongruent				
	8v9	0.037	0.41	0.287	0.499	***
		Congruent <<< Incongruent				

Interaction** Pairs x Device

		Pairs						
		1v2		2v8		8v9		F
Device	Computer	-0.179	0.4	-0.044	0.379	0.283	0.452	***
		1v2 < 2v8 <<< 8v9, 1v2 <<< 8v9						
	Tablet	-0.022	0.384	-0.048	0.381	0.082	0.413	*
		No significant pairwises						
	Smartphone	-0.051	0.359	-0.046	0.356	0.112	0.463	***
		1v2 << 8v9, 2v8 << 8v9						

Movement Time (z-scored)

Interaction** Pairs x Congruency

		Congruency				
		Congruent		Incongruent		F
Pairs	1v2	-0.217	0.38	-0.022	0.435	***
		Congruent <<< Incongruent				
	2v8	-0.152	0.384	-0.051	0.416	**
		Congruent <<< Incongruent				
	8v9	0.077	0.408	0.442	0.55	***
		Congruent <<< Incongruent				

Interaction** Congruency x Device

		Congruency				
		Congruent		Incongruent		F
Device	Computer	-0.046	0.407	0.06	0.44	**
		Congruent << Incongruent				
	Tablet	-0.12	0.361	0.146	0.455	***
		Congruent <<< Incongruent				
	Smartphone	-0.142	0.371	0.166	0.479	***
		Congruent <<< Incongruent				

Note: table continued on next page

Table A.3.1 *(continued)*

Interaction** Presentation Side x Device						
		Presentation Side				
		Left		Right		F
Device	Computer	0.008	0.43	0.019	0.417	ns
	Tablet	0.16	0.414	-0.134	0.402	***
		Left >>> Right				
	Smartphone	0.137	0.428	-0.113	0.422	***
		Left >>> Right				
Maximum Absolute Deviation (z-scored)						
Interaction** Pairs x Congruency						
		Congruency				
		Congruent		Incongruent		F
Pairs	1v2	-0.202	0.477	-0.037	0.526	***
		Congruent <<< Incongruent				
	2v8	-0.195	0.482	0.004	0.505	***
		Congruent <<< Incongruent				
	8v9	-0.065	0.55	0.56	0.672	***
		Congruent <<< Incongruent				
Interaction** Presentation Side x Device						
		Presentation Side				
		Left		Right		F
Device	Computer	0.049	0.427	-0.023	0.436	ns
	Tablet	-0.222	0.567	0.244	0.558	***
		Left >>> Right				
	Smartphone	-0.253	0.544	0.271	0.503	***
		Left >>> Right				

* 0.01, ** 0.001, *** 0.0001

Table A.3.2 Table of Sentence Verification means

Reaction Time (z-scored)								
Interaction*** Truth x Polarity x Device								
		Interaction*** Truth x Polarity						
Device	Computer			Polarity				
				Positive		Negative		F
		Truth Value	True	-0.592	0.213	0.559	0.257	***
				Positive <<< Negative				
			False	-0.205	0.183	0.303	0.21	***
				Positive <<< Negative				
	Tablet	Interaction*** Truth x Polarity						
				Polarity				
				Positive		Negative		F
		Truth Value	True	-0.417	0.26	0.404	0.298	***
				Positive <<< Negative				

Note: table continued on next page

Table A.3.2 *(continued)*

			False	-0.19	0.192	0.249	0.225	***
				Positive <<< Negative				
	Smartphone	Interaction*** Truth x Polarity						
				Polarity				
				Positive		Negative		F
		Truth Value	True	-0.418	0.255	0.397	0.317	***
				Positive <<< Negative				
			False	-0.18	0.192	0.238	0.183	***
				Positive <<< Negative				
Movement Time (z-scored)								
Interaction*** Truth x Polarity								
				Polarity				
				Positive		Negative		F
Truth Value		True		-0.319	0.258	0.416	0.332	***
				Positive <<< Negative				
		False		-0.174	0.215	0.105	0.266	***
				Positive <<< Negative				
Interaction*** Polarity x Device								
				Polarity				
				Positive		Negative		F
Device		Computer		-0.161	0.228	0.176	0.311	***
				Positive <<< Negative				
		Tablet		-0.276	0.245	0.283	0.292	***
				Positive <<< Negative				
		Smartphone		-0.308	0.207	0.329	0.263	***
				Positive <<< Negative				
Maximum Absolute Deviation (z-scored)								
Interaction*** Truth x Polarity								
				Polarity				
				Positive		Negative		F
Truth Value		True		-0.315	0.423	0.071	0.518	***
				Positive <<< Negative				
		False		0.112	0.421	0.15	0.389	ns
Interaction*** Truth x Device								
				Truth				
				True		False		F
Device		Computer		0.055	0.322	-0.043	0.278	ns
		Tablet		-0.191	0.514	0.196	0.448	***
				True >>> False				
		Smartphone		-0.241	0.493	0.253	0.413	***
				True >>> False				

* 0.01, ** 0.001, *** 0.0001

Table A.3.3 Table of Photo Preference High-Chosen means

Movement Time (z-scored)						
Main Effect: Pairs						
Pairs						
High-High		High-Med		High-Low		F
-0.045	0.425	-0.191	0.409	-0.282	0.744	***
High-High >> High-Med, High-High >>> High-Low						
Maximum Absolute Deviation (pixels)						
Main Effect: Pairs						
Pairs						
High-High		High-Med		High-Low		F
0.037	0.588	-0.084	0.507	-0.139	0.515	***
High-High >> High-Med, High-High >>> High-Low						

* 0.01, ** 0.001, *** 0.0001

Table A.3.4 Table of Photo Preference Matched-Pair means

Reaction Time (z-scored)								
Interaction** Reach Direction x Pairs								
		Pairs						
		High-High		Med-Med		Low-Low		F
Reach Direction	Left	-0.135	0.416	0.116	0.422	0.347	0.55	***
		High-High << Med-Med < Low-Low, High-High <<< Low-Low						
	Right	0.009	0.481	0.134	0.394	0.039	0.428	ns

Movement Time (z-scored)						
Main Effect: Pairs						
Pairs						
High-High		High-Med		High-Low		F
-0.045	0.425	0.117	0.442	0.269	0.513	**
High-High << Med-Med < Low-Low, High-High <<< Low-Low						

Interaction** Reach Direction x Device						
		Reach Direction				
		Left		Right		F
Device	Computer	0.096	0.409	0.032	0.476	ns
	Tablet	0.137	0.404	0.123	0.503	ns
	Smartphone	0.353	0.206	-0.045	0.412	***
		Left >>> Right				

* 0.01, ** 0.001, *** 0.0001