# Deep Learning-Based Behavioral Quantification of Upper Limb Rehabilitation Dose in a Rat Model of Ischemic Stroke

**Zanna Vanterpool**

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Science in Neuroscience

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa

## ABSTRACT

Seventy percent of stroke survivors experience loss of upper limb function after stroke and rehabilitative therapy is the only option to reduce impairments. However, uncertainty remains as to the optimal dose of therapy that should be prescribed. It has been suggested to report multiple parameters of dose, to increase standardization within the field, and to gain a better understanding of the dose-response relationship. This study investigated the automatic quantification of multiple dose parameters in a rat model of ischemic stroke, with rehabilitation paradigms whereby rats repeatedly grasp for food pellets to train their forelimb function. Starting 7 days post-stroke, groups of rats received 4, 8, or 12 rehabilitative training sessions for 10 days, practicing either high-quality (precision practice) or less skilled (mass practice) reaching movements. Pellet consumption was measured after each session and various metrics were analyzed using deep learning-based software (DeepLabCut, DLC) to represent parameters of dose intensity (number of reaches, paw path length) and session density (time on task). Functional outcome was assessed with the Montoya staircase task. Computer algorithms were validated against human analysis, demonstrating reach detection accuracy and reliability >80%. Interestingly, the number of training sessions did not alter the accumulated movement practice across rehabilitation, in either task. However, the number of sessions inversely affected training intensity, resulting in more forelimb use per session in rats with 4 sessions compared to 12 sessions. We found strong positive correlations between the number of reaches, time on task, paw path length, and pellets consumed in the precision practice, but only between reaches and pellets consumed in mass practice. This work demonstrates the quantification of multiple dose parameters using deep learning software and shows subtle differences between the two commonly used forelimb training tasks. Moreover, our data suggest that rehabilitative training at a frequency that is too high may negatively impact performance per session.

**TABLE OF CONTENTS**

## LIST OF FIGURES



## List of Tables



## List of Supplemental Figures

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AI | Artificial Intelligence |
| BDNF | Brain-derived neurotrophic factor |
| CIMT | Constraint-Induced Movement Therapy |
| DLC | DeepLabCut |
| EE | Enriched Environment |
| ET-1 | Endothelin-1 |
| FPS | Frames Per Second |
| GPU | Graphics Processing Unit |
| M1 | Primary Motor Cortex |
| MCAO | Middle Cerebral Artery Occlusion |
| MRI | Magnetic Resonance Imaging |
| NPV | Negative Predictive Value |
| PPV | Positive Predictive Value |
| PT | Photothrombosis |
| ROI | Region of Interest |
| SD | Standard Deviation |
| tPA | Tissue Plasminogen Activator |
| SMC | Sensorimotor Cortex |
| SRRR | Stroke Recovery and Rehabilitation Roundtable |
| STAIR | Stroke Therapy Academic Industry Roundtable |

**ACKNOWLEDGEMENTS**

First, I would like to thank Dr. Greg Silasi for his guidance and supervision throughout this project. I'm grateful for the opportunity I was given to join his lab and the experience I have gained in several areas. Next, I would like to thank Dr. Dale Corbett and Dr. Kate Hayward who collaborated with Dr. Silasi and conceptualized this project. Our monthly meetings ensured I remained on track and your input and advice were indispensable. I would also like to thank my advisory committee, Dr. Simon Chen and Dr. Diane Lagace for taking the time to provide feedback and stimulate critical thinking. I also want to highlight and express thanks to the ACVS staff for their tireless efforts and adaptability, especially through the pandemic.

Dr. Matthew McDonald and Matthew Jeffers, you have helped so many students with their projects over the years. I am extremely grateful for your assistance and fruitful discussions, and I would also like to acknowledge your contribution to the conceptualization of this project. Additionally, I would like to thank all current and previous lab members: Sarah Zhang, Melissa Filadelfi, Jungzheng Wu, Gavin Heidenreich, Célia Boumghar, Rana Abdelhalim, Katie Neale, Eilia Eslami Dizgah, and Zachary Eckert. Célia, you have put in tremendous effort analyzing videos by hand and it was essential to set up the automated analysis. Eilia, thank you for scoring additional data on short notice as well. I would also like to thank students from neighboring labs. Damian Chwastek, Candice Lee, Nima Raman, Sébastien Denizé, the social interactions outside of the lab made me feel welcomed when I first arrived.

I would like to thank Prins Bernhardt Cultuurfonds, Mullerfonds, and Fundatie van Renswoude for financially supporting my whish to continue my education abroad.

Finally, I would like to thank my personal support system. Mom, Dad, Axel, Alhassan, and members of team Phast, thanks for your advice, encouragement, and unconditional support. Without you, I would not have been able to complete this journey.

## STATEMENT OF CONTRIBUTIONS

**Junzheng Wu** designed and programmed all electronics, including the Pi camera recording setup

**Matthew Jeffers, Matthew McDonald, and Kate Hayward** designed the experimental set up for the mass practice experiment

**Matthew Jeffers** performed baseline training in the mass practice cohort

**Anthony Carter, Matthew Jeffers, Matthew McDonald** completed stroke surgeries in the mass practice cohort

**Matthew Jeffers, Matthew McDonald, and Zanna Vanterpool** completed the mass practice rehabilitative training paradigm

**Junzheng Wu** created the pellet presenter device including the assembly of the electronics and programming of the Arduino board.

**Zanna Vanterpool** assembled the training units in the precision practice experiment and completed pilot studies

**Zanna Vanterpool and Matthew McDonald** conducted the baseline training of the precision practice cohort

**Zanna Vanterpool, Matthew McDonald, and Matthew Jeffers** completed the precision practice rehabilitative training paradigm

**Zanna Vanterpool** wrote formulas for behavioral quantification within Excel

**Célia Boumghar** performed manual video tracking of the mass practice task

**Gavin Heindenreich and Zanna Vanterpool** collaborated in creating a Python script to process DeepLabCut output using the formulas written in Excel

**Zanna Vanterpool** was responsible for the development of the video analysis pipeline and trained two neural networks in DeepLabcut

**Zanna Vanterpool** performed all data analyses

# 1. INTRODUCTION

## 1.1 Stroke and stroke recovery in humans

Stroke is a debilitating medical condition in which hampered blood flow to the brain results in cell death and degeneration. Disrupted blood flow can either be caused by the rupture (hemorrhage) or occlusion (ischemia) of a blood vessel. Most strokes are ischemic in nature, accounting for 80% of cases (Truelsen, Begg, and Mathers 2006). There have been medical advances in acute stroke care, such as thrombolysis and thrombectomy. Thrombolysis involves treatment with tissue plasminogen activator (tPA), which is the only approved drug to dissolve the blood clot. It activates the production of plasmin to breakdown the fibrin that holds the clot together. However, treatment must occur within a short time window (4.5 hours after stroke), leaving many patients ineligible for treatment. Endovascular thrombectomy has a larger treatment window, 6-24 hours, however, only about 10% of ischemic stroke patients are eligible (Jadhav et al. 2018). Both forms of these acute stroke interventions have led to a reduction in mortality rate (Krueger et al. 2015). However, despite the improvements in acute treatment and reduction in mortality rate, functional outcome for those that survive a cerebrovascular attack is still poor. This has transitioned stroke from a disease of survivability to one of chronic disability.

It was estimated, in 2015, that 405,000 Canadians are living with a stroke-induced disability. Moreover, due to the aging population and other factors, this number is expected to increase to 725,000 by 2038 (Krueger et al. 2015). As such, the capacity of stroke survivors to perform most activities of daily living is greatly impaired, placing a significant burden on family members and the health care system. Although stroke results in a range of physical and cognitive disabilities, deficits of arm and hand function are most common, affecting 70% to 80% of stroke survivors (Nakayama et al. 1994; Broeks et al. 1999). As access to acute treatment options is limited, physical rehabilitation is the most promising intervention to facilitate the return of arm and hand function.

Although the mechanism by which rehabilitation improves function is not entirely known, multiple studies implicate brain plasticity. Several forms of brain plasticity exist, however, most generally it refers to the process of altering neural circuits due to environmental or internal demands and thus creating functionally appropriate neural connections (Kleim and Jones 2008). Naturally, neural plasticity can be induced by motor skill acquisition (Ward 2004; Plautz et al. 2000; Monfils, Plautz, and Kleim 2005; Rioult-Pedotti, Friedman, and Donoghue 2000). In this process, synaptic connections are persistently strengthened or weakened, resulting in long-lasting changes in signal transmission between neurons (i.e. long-term potentiation and depression, respectively). It is thought that these synaptic changes alter cortical circuitry, which can manifest in reorganization of movement representation (Adkins et al. 2006; Roman Siebner and Rothwell 2003). Therefore, physical rehabilitative interventions are fundamentally based on principles of motor learning. The Canadian Stroke Best Practice recommends task-specific training as a therapy for management of the upper extremity following stroke (Heart and Stroke Foundation of Canada 2019). This involves "the repeated practice of functional tasks, which combines the elements of intensity of practice and functional relevance. The tasks should be challenging and progressively adapted and should involve active participation" (Heart and Stroke Foundation of Canada 2019). Besides motor training-induced plasticity, there is evidence suggesting that ischemia creates a unique, transient, environment enhancing plasticity to a degree not seen with motor training alone (Zeiler and Krakauer 2013). For example, there is an upregulation of growth-promoting genes, increased rate of dendritic spine formation, and increased axonal sprouting (Carmichael et al. 2005; Brown et al. 2007; Carmichael and Chesselet 2002; Clarkson et al. 2013). The combination of spontaneous recovery, enhanced plasticity, and rehabilitative therapy results in the recovery profile that occurs within the first 3 months after stroke (Kwakkel, Kollen, and Twisk 2006; Langhorne, Bernhardt, and Kwakkel 2011).

Great effort is put into investigating various interventions that could further enhance motor recovery. However, a transformational change in stroke recovery, as we saw with thrombolysis and thrombectomy on the acute outcome, is yet to come. Multiple facets hamper progression in the field of stroke recovery and rehabilitation (Bernhardt et al. 2016), such as poorly defined interventions and lack of agreed methods for developing, monitoring, evaluating, and reporting interventions. Furthermore, there are gaps of knowledge pertaining to the dose and timing of motor rehabilitation. Although there is growing consensus that increased rehabilitation dose might be better (Lang, Lohse, and Birkenmeier 2015), there are still questions regarding the optimal dose and timing of rehabilitation. This is partly because rehabilitation dose is poorly defined, and therefore it has not been well quantified or controlled for (Lang, Lohse, and Birkenmeier 2015). Pharmacological dose, however, describes an amount of an active ingredient that will have the desired effect, taken at a specific frequency and duration (Lang, Lohse, and Birkenmeier 2015). Although task-specific training is known to be one of the key active ingredients in stroke rehabilitation (Kleim and Jones 2008; Nudo 2013), the precise working mechanism of action and timing thereof are not fully understood. Recently, a framework has been suggested to address the dose articulation problem within stroke recovery research (Hayward et al. 2021). This framework approaches dose as multidimensional. Meaning, that dose is "comprised of a duration of days that contain individual sessions and episodes that can be active (time on task) or inactive (time off task), and each individual episode can be made up of information about length, intensity, and difficulty" (Hayward et al. 2021). Intensity refers to how much of the task is performed per episode (work) or unit of time (rate). For example, 10 repetitions, 2 repetitions per minute.

To advance the stroke rehabilitation and recovery field, experts from across the globe have come together to establish guidelines on conducting and reporting research, including the pre-clinical recovery research (Bernhardt et al. 2016). The Stroke Recovery and Rehabilitation Roundtable (SRRR) advocates for better translation of pre-clinical evidence to human discovery trials in a bidirectional and iterative manner.

### 1.2 Animal models in stroke research

Animal models of stroke are essential for drug development and understanding the complex cellular and molecular pathophysiology, however, they may also contribute to our understanding of the role of neural plasticity following stroke (Murphy and Corbett 2009). Models of cerebral ischemia have been established in several species including rodents, pigs, cats, and non-human primates (Bacigaluppi, Comi, and Hermann 2010). The mouse is the most appropriate animal for genetic modification, making these animals suitable for transgenic studies on the molecular pathophysiology of stroke (Fluri, Schuhmann, and Kleinschnitz 2015). Rats, however, allow easier surgical manipulation because of their larger brain. Furthermore, compared to other species, rodents have a higher reproduction rate and a lower maintenance cost (Macrae 2011). Addressing clinical rehabilitation questions in preclinical research has several advantages compared to clinical trials. First, experimenters can control the gross lesion location (i.e. cortical vs. subcortical) and more precisely track the variability with accurate lesion volume measurements. This facilitates the creation of a homogeneous sample population. Second, higher sample sizes can be attained more easily. Third, it is possible to have within-subject baseline measurements and animals can be followed longitudinally in the experimental setting over time. Together, this facilitates systematic investigation of distinct interventions, within a reduced time frame. A downside of preclinical studies is that in many cases young, healthy male animals are used, which is not a good representation of the human stroke demographic. Stroke survivors are often middle-aged or older individuals with other health conditions such as: diabetes, hypertension, hyperglycemia, and atherosclerosis.

There are various ways to induce ischemic stroke in rodents. Middle cerebral artery occlusion (MCAO), which involves permanently or temporarily restricting blood flow through the MCA is most widely used. This artery is most often affected in human stroke. Blood flow to the MCA can be disrupted by an intraluminal suture inserted in the common carotid artery and advanced through the internal carotid artery, until it interrupts blood supply to the MCA (Fluri, Schuhmann, and Kleinschnitz 2015). The duration of the occlusion, ranging from 60-120 minutes, determines the extent of the damage, usually occurring in the striatum, frontoparietal and temporal cortices, as well as some portion of the occipital cortex (Garcia, Liu, and Ho 1995). A different method of transient ischemia is applying the potent vasoconstrictive peptide endothelin-1 onto an exposed vessel (e.g. MCA). It can be administered through injections, using stereotactic coordinates or onto the cortical surface. The severity of the lesion depends on the concentration of ET-1 that is used. ET-1 stroke induction is less invasive, has a low mortality rate, and can induce ischemia in deep and superficial brain regions. The photothrombosis (PT) model, also allows for targeted injury location through stereotactic coordinates (Watson et al. 1985). Ischemia is induced after a photosensitive dye (Rose Bengal) is injected (intraperitoneal or tail vein) and irradiated through the intact skull upon circulation. The activation of Rose Bengal by green light generates oxygen radicals that damage the endothelium and causes platelet aggregation. This is a minimally invasive procedure that can create highly reproducible strokes, however, profound edema and lack of reperfusion make this model different from human stroke (Labat-Gest and Tomasi 2013).

## 1.3 Animal behavior in stroke rehabilitation

### *1.3.1 Why is behavior important?*

There are many tools available for detailed examination of brain structure and neurophysiology following stroke, such as two-photon imaging and optogenetics. However, a general aim of neuroscientists is to understand brain functions, such as cognition, memory, and motor control, therefore studying only parts of the brain will not be sufficient to understand how these functions generate behavior (Krakauer et al. 2017). Behavior is important because it is ultimately the output of many brain functions (Anderson and Perona 2014) and it is a way of interacting with the world around us. Previously, behavior has been defined as ‘‘The total movements made by the intact animal’’ (Tinbergen 1955). But a more recent definition states “Behavior is the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes’’ (Levitis, Lidicker, and Freund 2009). As such, behavior allows organisms to quickly adapt to a changing environment. Therefore, to understand brain function, it should be considered in the context of behavior (Anderson and Perona 2014; Krakauer et al. 2017).

Pre-clinical stroke researchers examine animal behavior to understand how brain function is altered by ischemia and can be restored by subsequent therapeutic interventions. These approaches range from brain stimulation techniques and stem cell therapies, to pharmacological rehabilitation and movement therapy. For the scope of this thesis, I will focus on behavioral movement therapy. Over the last several decades, basic neuroscience has taught us many principles of experience-dependent neural plasticity (Kleim and Jones 2008; Nudo 2013). For example, specificity, repetition, intensity, and timing of training all matter. Studies have shown that plasticity is specific to the type of experience. Changes in patterns of neural activity and motor maps require the acquisition of a motor skill, and will not emerge following simple motor activity alone or repetition of previously acquired motor movements (Plautz et al. 2000). For example, rats performing skilled reaching movements demonstrate an expansion of the cortical forelimb representation, but unskilled movements do not produce this effect (Kleim et al. 2002; Remple et al. 2001). Motor skill acquisition leads to dendritic growth, synapse addition, and gene expression (Monfils, Plautz, and Kleim 2005; Nudo 2003; Rioult-Pedotti, Friedman, and Donoghue 2000). Furthermore, newly learned behaviors should be repeated to induce long-lasting changes. Improvements in skill may precede structural changes and motor map reorganization by several days (Monfils, Plautz, and

Kleim 2005). Thus, plasticity through repetition can be viewed as the concretization of the new skill, making it more resistant to decay in the absence of training (Monfils, Plautz, and Kleim 2005). In addition to repetition, to produce lasting changes within the brain, the intensity of behavioral training must be sufficiently high. Rats performing 250 reaches per day showed enhanced recovery from ischemia together with increased levels of brain-derived neurotrophic factor (BDNF) compared to animals performing 150 reaches (MacLellan et al. 2011). Furthermore, transcranial magnetic stimulation of the primary motor cortex (M1) in humans can elicit a long-lasting increase in motor evoked potentials with stimulation trains of 1800 pulses but not 150 pulses (Peinemann et al. 2004). However, training intensity can also have an adverse effect on motor recovery. It has been shown that exclusive use of the affected limb (24/7) early after sensorimotor cortex lesions exacerbates functional outcomes (Humm et al. 1998). This also demonstrates that the timing of rehabilitative training matters. There are cascades of neural reactions following brain injury that occur over several months, e.g. upregulation of growth-promoting and inhibition factors, (Murphy and Corbett 2009). One thing to consider is whether the behavioral therapy is intended to have a neuroprotective effect, sparing loss of neural tissue, or a restorative effect, inducing cortical reorganization (Kleim and Jones 2008). Most rehabilitative therapies promote restructuring of neural circuits. Although cortical reorganization is possible at any point in time, there is a window of ischemia-induced heightened plasticity, as mentioned above. Animal studies indeed suggest rehabilitative training is more effective when delivered days after stroke, compared to weeks or months (Biernaskie, Chernenko, and Corbett 2004; Barbay et al. 2006).

### *1.3.2 What kind of behavior is studied?*

Naturally, rodents exhibit a wide variety of instinctive behaviors linked to survival. One can think of reproductive behavior, aggressive behavior, locomotion, and feeding behavior, for example. Experimental stroke recovery research mostly studies sensorimotor behaviors. These types of behaviors involve skilled hand use, locomotion, and asymmetry. There is a wide variety of assays available to capture (parts of) complex behavior (Schaar, Brenneman, and Savitz 2010; Ruan and Yao 2020). However, it is up to the experimenter to choose tests that are sensitive to the induced brain damage and applied intervention (Schaar, Brenneman, and Savitz 2010). Although a lot of behavioral paradigms leverage the natural occurrence of elements of behavior (e.g. reaching-and-grasping motions; exploration; olfaction; gait), in an experimental setting animals must typically learn to perform the paradigm. Some paradigms are relatively simple and require little to no training. In the cylinder task for spontaneous forelimb use, rats or mice are placed in a transparent Plexiglas cylinder (20 x 30 cm for rats) and their exploratory behavior is recorded with a camera. Non-injured rodents will touch the wall and land on either paw equally. Impaired animals, however, will show limb-use preference towards the healthy paw (Schallert et al. 2000). Slightly more complex tasks, like ledged tapered beam walking, require 2-3 days of training. This test assesses fore- and hindlimb locomotion as animals walk across an elevated tapered beam with an under-hanging ledge, and the number of foot faults is quantified (Schallert 2006). Following stroke, animals show increased foot faults on the contralateral side, as the beam tapers. Numerous behavioral paradigms evaluate skilled hand use in rodents. This group of tests elicits a complex sequence of movements and requires a substantial amount of time to shape the behavior (i.e. 2-4 weeks) (Schaar, Brenneman, and Savitz 2010). Generally, rats and mice are trained to perform trials in which they grasp a food object (i.e. a pellet, seed, or pasta), which will be further discussed in chapter 1.4. Skilled reaching tasks are recommended assessments due to the ability to detect long-lasting impairments, more than 4 weeks after stroke, and sensitivity to lesion size (Corbett et al. 2017). Importantly, the similarities between human and rodent reaching movements give these tasks a high translational value (Klein et al. 2012; Lieshout et al. 2021). However, one should keep in mind that rodents use olfaction to localize the pellet (Whishaw and Tomie 1989) whereas humans use vision. Also, the sugar pellets that are often employed as the target for rodent reaching tasks are highly rewarding, while humans have less incentive during reaching tasks. Finally, since behavioral assays target specific modalities of function, it is recommended to use a battery of tests spanning a range of sensory-motor functions similar to those used in human stroke (Corbett et al. 2017).

#### *1.3.3. How is behavior analyzed?*

Traditionally, biologists assessed animal behavior qualitatively through detailed (manually recorded) descriptions. Over the last 40-50 years, however, this has moved more towards scoring behavior according to certain criteria, especially in the laboratory environment (Anderson and Perona 2014). Presently, we discern three main ways of evaluating behavior: end-point measures, movement description, and kinematics (Whishaw et al. 1999). End-point measures are the measures of the final outcomes of an action, e.g. a lever was pressed, a pellet was eaten, or an object was touched. How this action occurs can be specified using systems that use formal languages for movement description, such as the Eshkol Wachman Movement Notation (Eshkol and Wachman 1958). This system was used by Whishaw and colleagues to describe and rate movement components of a reach-to-eat task in rats (Whishaw and Pellis 1990). When actions are described as Cartesian representations, giving information about trajectories and angles between body parts they fall into the kinematics category (Whishaw et al. 1999). This is a mathematical and objective approach that is sometimes referred to as the "geometry of motion". Kinematic analysis requires video capture of the behavior, however, most modern cameras available today (even cell phone cameras) are suitable for this. For kinematic analysis of quick and short movements (such as reaching), it is advisable to use equipment that can film at a high frame rate (~ >60 frames/second) and a short shutter duration (<0.001 s) to obtain the best results (Whishaw et al. 1999). Following the behavioral assay, the videos are analyzed frame-by-frame by the experimenter for manual scoring. This is a lengthy process as human observation takes 3x the length of the video (Anderson 2014). Consequently, as an experimenter may become fatigued it is more likely that errors occur in the analysis. Furthermore, although there may be well-established protocols for video analysis within labs, the scoring of behavior by experimenters remains subjective. Kinematics are not subjective, but this type of behavioral analysis often requires the placement of markers for tracking by software packages, which animals might find intrusive. On a final note, it is important to be aware that behavioral assays are often conducted outside of the animal's home cage environment and during the daytime. Stressful transportation and the interruption of the circadian cycle could impact the behavior that is assessed (Koch et al. 2017).

## 1.4 Upper limb rehabilitation & recovery in rodent models of stroke.

### *1.4.1 Training and assessment of skilled upper limb function in rodents.*

Animal studies investigating upper limb function following stroke can use several tasks to assess impairment. As discussed in the previous section, some assays measure pre-existing sensorimotor behaviors, which are "unskilled" tasks (e.g. cylinder task, adhesive tape removal, foot fault tasks) (Kleim, Boychuk, and Adkins 2007). Others, measure acquired, "skilled", behaviors, which typically involve reaching for food pellets through a narrow space (e.g. a slit or staircase). This latter group will be briefly discussed here.

Apart from the substantial amount of training that skilled forelimb behaviors require, they usually also require food restriction up to ~90% of the baseline body weight. This is done to increase motivation to obtain the food rewards. In the Montoya staircase task, animals are trained to retrieve pellets that are placed in shallow wells of a double staircase (also see Methods 2.2) (Montoya et al. 1991). The staircases are separated by an elevated platform, that provides bodyweight support and in combination with a low ceiling, it prevents the animal from retrieving pellets with the contralateral arm. Contrary to the staircase task, obtaining pellets in the tray reaching task does not become progressively more difficult due to increased reaching distances. The animals are placed in clear boxes that are open at the front with small metal bars. Then, they learn to extend their paw between the bars to grasp pellets that are displayed on a long shelf outside of the box (Whishaw, O'connor, and Dunnett 1986). The single pellet reaching task (Whishaw et al. 1991), which is a widely employed behavioral test, is considerably more difficult than tray reaching. Now, animals must target a single pellet on a shelf rather than simply grasping from a tray full of pellets. Moreover, the pellet is placed off-center, such that the animal can only retrieve the pellet with the contralateral paw. To distinguish between trials, animals are trained to walk to the back end of the box, which also allows the experimenter to place new pellets.

Assessment in skilled reaching tasks is usually bound either by time, a maximum number of pellets, or both. For a testing session, parameters of end-point measures are most often reported, these include the percentage of successfully retrieved pellets, percentage of successful reaches, and the number of reaches. Other evaluation methods (i.e. movement description and kinematics) are implemented less frequently due to the significant time required to analyze.

*1.4.2 Preclinical upper limb rehabilitation strategies in rodents.*

Considering that behavioral experience is the most potent modulator of neuroplasticity it is often a key component in rehabilitation studies, either on its own or in combination with a novel therapeutic intervention. For instance, rehabilitation strategies targeting upper limb function comprise forced limb use, voluntary exercise, and reach training. Substantial return of function can be observed in 2-4 weeks post-stroke, even in animals with large cortical lesions (Alaverdashvili et al. 2008; Krakauer et al. 2012). These animals demonstrate recovery of success rate on end-point measure scales by compensatory movements (Alaverdashvili et al. 2008).

Animals with smaller lesions exhibit both restitution of pre-stroke movement patterns and compensatory patterns (Moon et al. 2009).

Constraining the use of the less impaired limb in rodents, by treating it with a plaster cast or bracelet, is a training strategy that was first applied by Wishaw (Whishaw, O'connor, and Dunnett 1986). It forces animals to use their impaired limb during rehabilitation. Studies employing this constraint-induced movement therapy (CIMT) demonstrate enhanced behavioral outcomes (Zhao et al. 2013; Nesin et al. 2019), mediated by enhanced synaptic plasticity and axonal growth. Nesin and colleagues (2019) immobilized the ipsilesional forelimb starting 5 days after ET-1, for 7 days and assessed the quality of skilled reaching and success rate. They demonstrated increased dendritic arborization in layers 2-3 of the injured motor cortex, compared to rats without therapy. Furthermore, CIMT was shown to be effective early after subcortical hemorrhagic stroke as well (Ishida et al. 2015).

Rehabilitative interventions investigating the effect of aerobic exercise on recovery outcomes have found mixed results. Al Shoyaib and colleagues (2021) demonstrated that when PT-induced mice had overnight access to a running wheel for 6 days per week, they showed improved recovery of motor function compared to mice with only 5 hours access. This was attributed to enhanced angiogenesis. However, the animals were assessed on a locomotion task and not skilled reaching. Others, however, suggest that voluntary exercise or strength training as an additional component of a reach training paradigm does not improve skilled reaching (Maldonado et al. 2008; Remple et al. 2001).

Conversely, environmental enrichment (EE) as an adjunctive therapy has shown to be effective in improving recovery outcomes (Johansson and Ohlsson 1996; Biernaskie and Corbett 2001). Although there is no standardized form of EE, it usually involves socially housing animals together in a large cage, supplemented with toys, ladders, ropes, and more, that are rearranged every couple of days. Which exact aspect of this multi-faceted form of housing (i.e. motor, cognitive, sensory, or social stimulation) is responsible for the improved outcomes is not fully understood. However, animals exposed to EE show increased growth-promoting factors, neurogenesis, vascular remodeling, axonal sprouting, and less white matter damage and growth-inhibiting factors resulting in improved cognitive and gross motor function (McDonald et al. 2018).

When it comes to rehabilitative paradigms that specifically employ reach training to enhance forelimb function, it was shown that, indeed, EE is an effective adjunctive therapy that can be referred to as enriched rehabilitation (Biernaskie and Corbett 2001; Jeffers and Corbett 2018). Animals receiving 6 hours of trough reach training during the dark cycle, combined with EE, demonstrated significant recovery, which was attributed to increased levels of BDNF and a higher number of pellets consumed (~250) (MacLellan et al. 2011). These data suggest that there is a critical threshold of training intensity above which rehabilitation becomes effective. It is unknown if such a threshold exists in humans. However, currently, stroke patients typically perform far fewer movement repetitions, with only ~32 in 1 hour therapy (Lang et al. 2009). Despite it being feasible to attain numbers that demonstrated efficacy in animal studies (Birkenmeier, Prager, and Lang 2010). In mice, a higher dose (2x training session) of skilled reach training was found to expedite the return of pre-stroke motor performance compared with a lower dose, that required more training sessions (Bell et al. 2015). One week of single pellet reach training (45 minutes, or 100 pellets daily) was also found to be effective in improving performance in an unrelated task, sticky tape removal (El Amki et al. 2017). Whereby animals that received rehabilitative training had a shorter latency time, suggesting transferability of skilled motor training to other tasks. Finally, Jeffers and colleagues (Jeffers et al. 2018) demonstrated that rehabilitation and enrichment is crucial for rat stroke recovery. They showed that biomarkers of initial impairment and infarct volume can be used in a model to prescribe the specific dose of daily rehabilitation required and thereby reliably predict recovery. This individualized approach was especially effective in rats with severe impairments.

Thus far, the majority of preclinical studies investigating physical rehabilitation to promote recovery record a measure of training intensity (pellets consumed) and/or the number of sessions (e.g., twice daily). This underscores the unidimensional reporting of dose which may hinder a deeper understanding of the active ingredients of rehabilitative training and their mechanism of action. Following the notion that rehabilitation dose is multidimensional and should be better articulated in preclinical studies as well, researchers should aim to also incorporate other dimensions. For example, for what duration of the training sessions is the animal “on task” and actively performing the training? This can also be expressed as a percentage, i.e., “session density”. Furthermore, using pellets consumed as a metric to indicate training intensity does not capture discrepancies in reaching strategies that animals portray, which could potentially play a role in the recovery outcome. Kinematic variables do have the capacity to reveal this. In addition, the dimension "sessions" has not yet been systematically investigated in rodents. Some studies in the Corbett lab (i.e., Biernaskie and Corbett 2001; Biernaskie, Chernenko, and Corbett 2004; Jeffers and Corbett 2018) gave rats access to the reaching apparatus several hours per day, which resulted in substantial pellet consumption. These results also support the understanding that "more is better", regarding pellet consumption and subsequent recovery. However, it is unknown (1) whether more training bouts per day are also associated with more pellet consumption and (2) how to efficiently deliver training sessions that elicit high-intensity forelimb use. A sophisticated examination of multiple dose metrics and (simultaneous) investigation of dose escalation is difficult to conduct using conventional means of manual, frame-by-frame, video analysis. Thus, addressing these gaps of knowledge warrants a different approach for behavioral analysis which will be discussed in the next section.

### 1.5 Advancements in behavior analysis.

Detailed analysis of full-length training sessions by using standard means of analysis takes prohibitively long. Perhaps this is the reason why no one has previously reported a study that undertook this challenge. Analyzing snippets of video recordings and inferring overall behavior from this, a method called subsampling could have been an alternative. However, often behavioral assays were not filmed (Biernaskie and Corbett 2001; Biernaskie, Chernenko, and Corbett 2004; Jeffers and Corbett 2018). Video recordings require a lot of costly storage space, although, with rapid technological developments, this has become much cheaper. Another important development that is currently shifting the approach to behavioral analysis is that of artificial intelligence (AI). AI is the intelligence with which machines, software, and devices independently solve problems. They imitate the human mind. The roots of AI can be traced back to the 1940s and was officially coined in 1956 (Haenlein and Kaplan 2019). However, it was not until the first year of the 21$^{st}$ century when AI underwent exponential growth. Nowadays, AI is fully integrated into our daily lives, for example through facial and speech recognition features on our phones, and self driving cars. More recently AI has also found its application in behavioral analysis.

Over the last two decades, increasing efforts have been directed to automation of behavioral analysis (Anderson and Perona 2014). Automated systems rely on instruments that can detect movement. Examples include accelerometers, infrared beam-breaks, and reflectors. However, video recording is the most commonly used modality due to its high spatial-temporal resolution (Anderson and Perona 2014). Essentially, software systems can be divided into three modules: tracking, action classification, and behavior analysis. Briefly, tracking involves computing the course of an animal's location, since this is the most popular modality it will be further discussed below. Action classification entails the identification of whether an action (e.g. grooming or walking) is performed. This analysis is more complex and requires training (i.e. machine learning) of the software to recognize specific actions (Kabra et al. 2013). Machine learning is a part of AI whereby algorithms learn to make predictions through pattern recognition. Ultimately, through automatic behavioral analysis, patterns of behavior consisting of multiple actions would be estimated (Anderson and Perona 2014).

Many software packages, either open-source or commercial, have become available in recent years that enable tracking of many species, including rodents (Pham et al. 2009; Gomez-Marin et al. 2012; Ohayon et al. 2013; Kabra et al. 2013; Ben-Shaul 2017). Distinguishing the animal from the environment is pivotal, in order to track it. Typically, this is based on the contrast between the animal and the background. Then, specific features outlining the body are identified, for example, the nose, tail base, tail tip (Ben-Shaul 2017). This is used to measure the orientation of the body, which is called pose. These pose positions are then concatenated frame by frame, to obtain a trajectory. Unfortunately, most software packages are restricted to specific species and behavioral paradigms. Furthermore, the underlying code is difficult to modify to an experimenter's needs. The drawback of a plethora of restricted systems is that it hampers generalizability and standardization. Biologists, ecologists, and neuroscientists would benefit from one package that can track virtually any organism across different settings.

Fortunately, a package that meets this need has recently been released and widely adapted within the neuroscience community. This tool, DeepLabCut™ (DLC) is freely available and estimates pose without the use of any kind of markers (Mathis et al. 2018). DLC uses deep learning to recognize user-defined features of animals. Deep learning is a form of machine learning that is more advanced because it uses a computing model inspired by the structure of the brain. As a result, deep learning algorithms can improve themselves with less human supervision. Because the DLC package is pre-trained on a large dataset of images (ImageNet), it only needs a few hundred training images provided by the experimenter to accurately extract those features of interest (Mathis et al. 2018). Subsequently, post-hoc analysis is performed on the estimated poses to extract behavior. It has been shown that DLC has an equal, if not better, performance in behavioral analysis as commercial packages (Sturman et al. 2020). Furthermore, the markerless tracking by DLC enables kinematic analysis without the use of intrusive markers (Bova et al. 2020; Weber et al. 2021; Moro et al. 2020). Tracked body parts no longer need to be determined a priori which makes it possible to capture more of the animal's behavior. Additionally, depending on the resolution and computational power, videos can be processed up to a speed of 450-600 frames per second. That is 15-20x faster than regular playback speed for a video recorded at 30 frames per second (Mathis and Warren 2018). This significantly reduces video analysis time and facilitates large-scale movement data analysis. Full-length video analysis can give us more insight into the distribution of natural behavior, for example, equal spread vs. a burst of activity across the duration of an assessment. Moreover, once a neural network is established it can repeatedly be used within labs and exchanged across labs, removing experimenter bias. Although training a network in DLC is designed to be "accessible", some labs might still face some challenges with post-processing DLC data if there is a lack of computer science knowledge. Together, however, DLC has the potential to change how we approach behavioral video analysis and the information we obtain from it.

## 1.6 Rationale

In conclusion, there is a need stemming from clinical stroke rehabilitation field to better report and understand rehabilitation dose. Preclinical research has been fundamental in establishing our current knowledge of many rehabilitation related principles. As such, animal models of stroke are useful to investigate basic principles specifically pertaining to rehabilitative training dose. In doing so, registration of behavior is important to capture different constructs of dose. However, such a sophisticated analysis, at a large scale, is practically impossible by traditional means of video registration and manual analysis. Fortunately, novel tools for movement registration enable high video throughput. This opens the door to using video-derived metrics of behavior in large quantities, to systematically investigate the effect of rehabilitation dose escalation on task engagement and forelimb use in two commonly used reaching tasks in rats following stroke.

**Hypothesis:**

DeepLabCut pose estimation can be used to assess components of upper limb rehabilitation dose that may relate to recovery outcome. In addition, the total amount of the dose components will increase (proportionately) with the number of rehabilitative training sessions.

**Objectives:**

1. To develop a pipeline that uses markerless tracking to automatically quantify time on task, number of reaching attempts, and paw path length
2. To determine what number of reaching sessions results in the most time on task, number of reaching attempts, and paw path length
3. To determine the relationship between video-derived metrics and pellet consumption

## 2. MATERIALS & METHODS

### 2.1 Animals and behavioral training

*2.1.1 Animals*

We used female Sprague-Dawley rats (N = 40, Charles River) weighing 200-225 grams and approximately 6 weeks of age on arrival. The rats were housed in groups of 4 in a reversed 12-hour day-night cycle. After acclimatization to the animal facility, rats were handled during their night cycle. The animals were food restricted up to 95% of their normal body weight during initial training and testing in the Montoya staircase task (Montoya et al. 1991). All experimental procedures were done in accordance with the Canadian Council of Animal Care guidelines and approved by the University of Ottawa Animal Care and Veterinary Service.

*2.1.2 Montoya staircase task*

Rat skilled forelimb performance was assessed using the Montoya staircase task. The apparatus consisted of a narrow rectangular Plexiglas chamber containing a double staircase in the front (7 steps each) separated by a plateau supporting the rat's body weight (**Figure 2A**). Rats reached for sugar pellets (45 mg; Test Diet; Purified Rodent Tablet; 5TUL) placed into shallow wells of each step (3 per step). Due to the narrow width and height of the chamber, the forelimbs cannot crossover the plateau, requiring ipsilateral forelimb use on each side of the Montoya task. Pellet retrieval becomes increasingly more difficult as steps lower. Rats were placed in the chamber for 15 minutes, after which the number of pellets retrieved were counted. Rats were trained on this task twice daily, for 10 consecutive days. Their baseline performance was measured on the last two days. For animals in the mass practice experiment, the paw with greater mean performance on the last two days was defined as the dominant paw. In the precision practice experiment, paw dominance was determined based on pellet retrieval in the single pellet reaching task described below. Recurring assessments, before (week 1 post-stroke) and after rehabilitation (week 3 post-stroke) comprised 3 testing days preceded by overnight food restriction (**Figure 1A**).

*2.1.3 Trough reaching task*

In the trough reaching task, rats were placed into similar boxes as to the staircase task, however, the staircases were replaced with a singular trough positioned beneath the dominant paw (**Figure 2C**). The trough was loaded with 10 grams of pellets to promote "unlimited" forelimb use. Pretraining took place for 3 days (5 sessions) after completion of the Montoya staircase test and prior to stroke induction (**Figure 2A**). A training session lasted 15 minutes. In session 1 the trough was placed at mid-height which allowed rats to lick pellets. Sessions 2-5 involved a low-height tray to reinforce forelimb use for pellet retrieval. The difference in weight of the pellet trough after a training session was used to calculate the number of pellets retrieved.

*2.1.4 Precision reaching task*

The precision reaching task was designed as a modified version of the single pellet reaching task as described by (Whishaw et al. 1991) (**Figure 1D**). Our task includes automated pellet presentation but does not require walking to the back end of the chamber between trials. Rats were placed in a Plexiglas box (35 x 13 x 38 cm) with a 1.3 cm wide and 10 cm tall slot in the front wall extending 3 cm from the floor surface. A food hopper was externally located in front of the reaching box and mounted onto the edge of the shelf on which the box is situated (**Figure 2C-F)**. Most hopper parts were 3D printed (FlashForge 3D printer, Creator Pro) and assembled in the lab. The device contained an Arduino Nano (Arduino, 7630049200173) that controlled a servo motor (Hitec, HS-82MG). This motor drove the pellet presenter arm, which cyclically displayed a pellet for 4 s in a small indentation on top of the arm (~120 presentations in 15 minutes). Pellets were aligned with either lateral edge of the slot, 2 cm forward and 1.5 cm upward from the bottom of the slot, allowing rats to use only left or right paw to obtain the food reward. The somewhat unstable positioning of the pellet necessitates skillful and directed movements to grab the pellet without displacing it into the hopper. After each session, pellet consumption was determined by measuring the difference in weight of the pellets stored in the hopper.

Baseline training in this task lasted 4 days (16 sessions, 15 minutes each). In sessions 1-2, rats freely grasped pellets displayed on a tray (**Figure 2F)**. Licking was discouraged by gradually moving the pellets further from the reaching slit. The pellet presenter apparatus was introduced in session 3, presenting pellets in the center of the slit. Experimenters observed paw preference and in sessions 4-16, pellets were presented in accordance with the preferred limb. This paw was determined as the dominant paw and targeted during stroke surgery. Rats with an average number of pellets retrieved <11 for sessions 5-16 were considered "non-reachers" and excluded (n = 5).

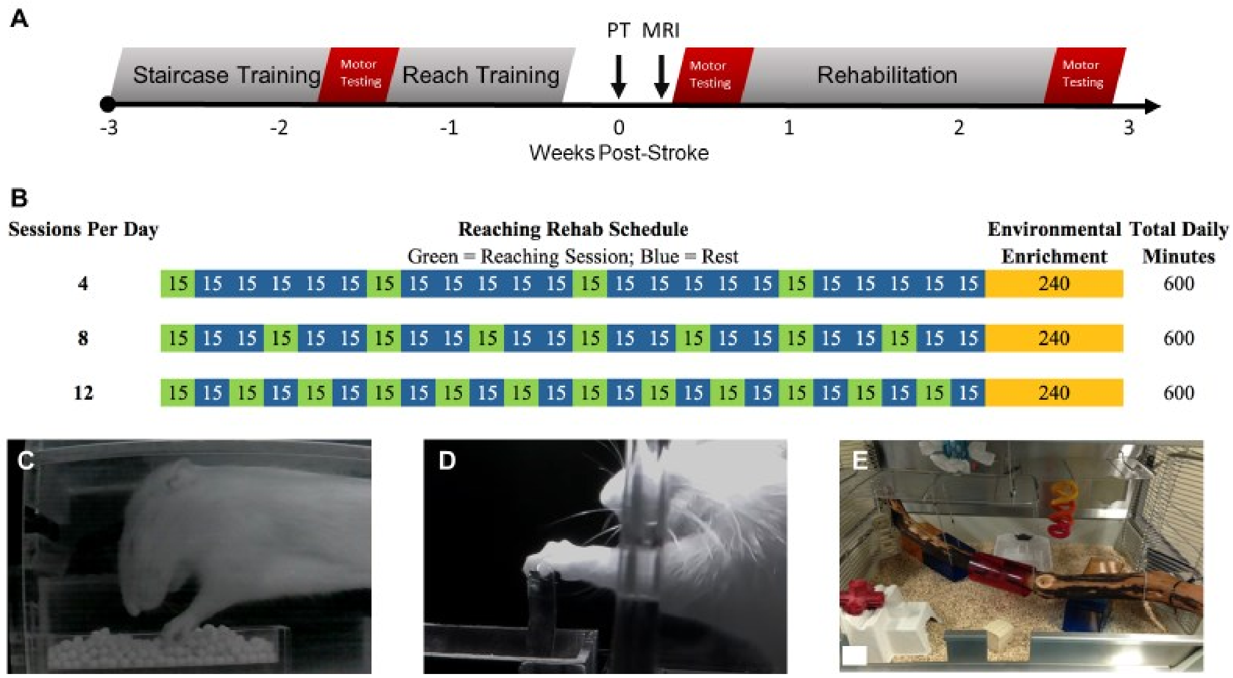


**Figure 1. Experimental overview. A)** Experimental timeline. **B)** Rehabilitation day schedule in minutes. Green blocks indicate 15-minute reaching sessions, blue blocks indicate 15-minute home cage resting sessions. **C)** Rat reaching in the mass practice rehabilitation whereby pellets are presented in a trough. **D)** Rat reaching in precision practice task whereby the rat aims for the pellet on top of the pedestal through the slot in the front wall. **E)** Enriched environment cage with toys, shelter, and tubes. PT = photothrombotic stroke, MRI = magnetic resonance imaging.

## 2.2 Photothrombotic stroke

Stroke was induced with cold-light photothrombosis using a 1 cm circular diameter halogen lamp (Intralux 5100, Harvard Apparatus). Rats were anesthetized with isoflurane (5% induction, 2% maintenance in 100% $O_2$ given at 1.6L/min) and body temperature was maintained at 37°C using a heating blanket. Saline (1 mL) was administered subcutaneously to prevent dehydration during surgery. The dorsal surface area of the head was shaved and wiped with an antiseptic (SoluPrep, 3M), and the animal was placed in a nose cone. Then, a sagittal midline incision in the scalp exposed the skull overlying M1. Next, an aluminum foil illumination aperture was aligned to Bregma and the midline of the skull, with the light source placed over the skull at +2.3 mm anterior and ±2.5 mm lateral to Bregma. The hemisphere contralateral to dominant paw during the Montoya staircase task (mass practice experiment) or single pellet reach training (precision practice experiment) was targeted for focal ischemia. The photosensitive dye Rose Bengal (20 mg/kg in

0.9% NaCl solution), was injected into the rats' tail vein. After 2 minutes of circulation in the vasculature, the cold light source was turned on and irradiated the skull for 10 minutes. Finally, the incision was sutured, and 0.2 ml of bupivacaine (Chiron) was applied along the sutures as a topical analgesic. The animals recovered from anesthesia in a heated chamber before returning to their home cage. Post-operative health checks were conducted at 4 hours (including reapplication of bupivacaine) and the morning following surgery. One animal in the mass practice cohort died following surgery.

### 2.3 Infarct volume quantification

The presence of infarction was confirmed using magnetic resonance imaging (MRI) at 48 hours post-surgery. Animals were anesthetized by inhaling isoflurane (5% induction, 2% maintenance in 100% $O_2$ given at 1.6L/min). Then, T2-weighted structural images were obtained using the MRI bore (7T General Electric/Agilent MR901 small-animal scanner). The parameters were the following: 21 coronal slices; slice thickness = 800 mm; in-plane resolution = 132.8 mm; echo train = 8; echo time = 27 ms; scan time = 5 min. The images were analyzed in ImageJ (National Institute of Health) to determine lesion volume and extension.

### 2.4 Rehabilitation paradigms

The two reaching tasks described above were employed in two rehabilitation regimens. In the mass practice paradigm, animals performed the trough reaching task and in the precision practice paradigm, animals were engaged in the precision reaching task. Rats were divided into three subgroups receiving 4 (n = 6 mass practice; n = 5 precision practice), 8 (n = 6 mass practice; n = 5 precision practice), or 12 (n = 6 mass practice; n = 5 precision practice) reaching sessions per day. Each reaching session was 15 minutes long and spaced equally throughout the day with resting sessions (15-75 minutes) in their home cage (**Figure 1B**). Following completion of reach training, animals were moved into an EE for 4 h. Ferret cages (~ 81 x 82 x 60 cm) were arranged with toys and ramps to encourage explorative behavior and physical activity and rats were housed in groups of 4. The total daily rehabilitation schedule was 10 hours. Rehabilitative training was administered

5 days per week (Mon-Fri), for 2 weeks. In the mass practice experiment, rehabilitative training lasted from day 7-20 post-stroke and in precision practice from day 5-16 post-stroke.

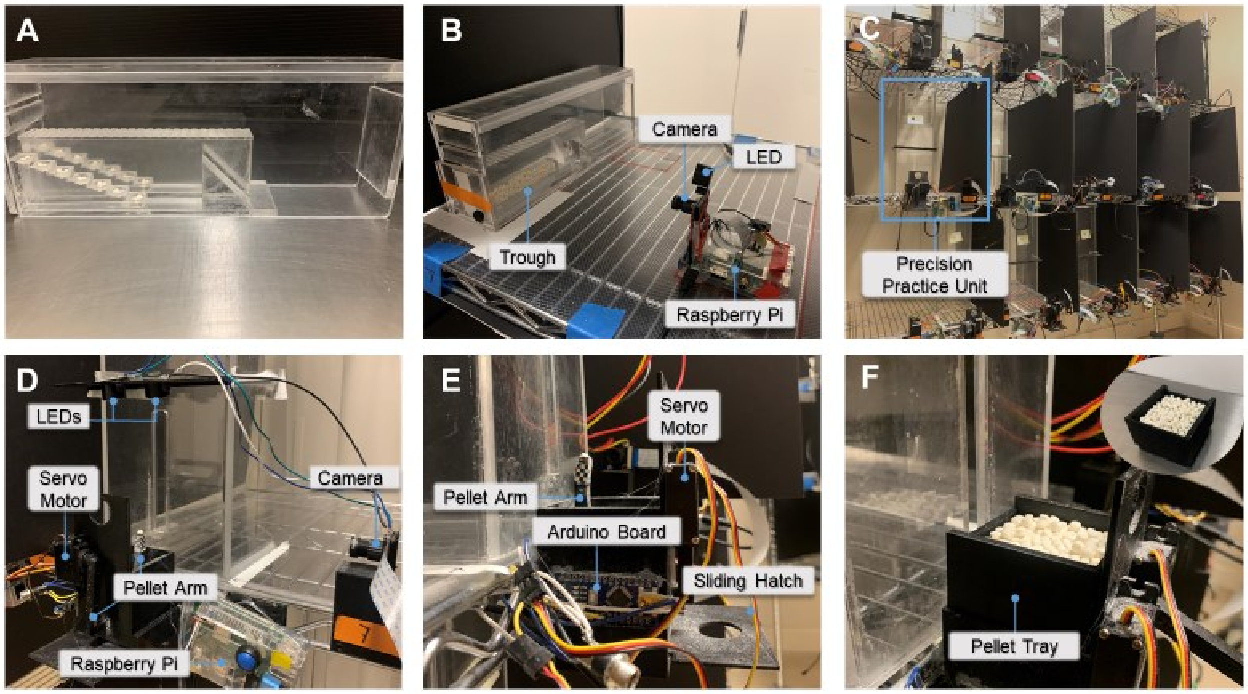


**Figure 2. Skilled reaching apparatuses. A)** Staircase apparatus with pellets in wells. **B)** Setup of a single unit of the mass practice rehabilitation. **C)** Rack containing all (15) precision practice units; a single unit is outlined. **D)** Front view of precision practice unit. Custom parts were 3D printed to attach the camera, Raspberry Pi, and pellet hopper to the front edge of the shelf and the LEDs to the reaching chamber. **E)** Side view of pellet hopper. Opening the sliding hatch releases the pellets, allowing them to be weighed after collection, without having to remove the entire device. **F)** Pellet tray that was used during trials 1-2 of precision reach training. The tray can be placed on top of the hopper at the same height as single pellets were presented.

### 2.5 Video acquisition

All rehabilitative reaching sessions were filmed with infrared-sensitive Raspberry Pi cameras (Waveshare: SKU: 851-1, UPC: 799632838265) at 30 frames per second (fps) with a resolution of 1296 x 760 pixels and a shutter time of 3.5 ms (mass practice) or 1.0 ms (precision practice). The cameras were aligned with the sagittal plane of the rat's body, ipsilateral to the reaching arm. Infrared LEDs provided illumination of the reaching boxes to enable recording during the dark cycle. We used a custom python script to automatically store video recordings in a Google Drive folder.

## 2.6 Video analysis

Once all behavioral testing was completed, we started building a pipeline to analyze the videos. First, we investigated the feasibility of using algorithms to characterize reaching behavior on manually tracked video samples. Second, we explored ways of reducing analysis time and retaining accuracy. This was gauged in mass practice videos through visual analysis and subsequently was applied in multiple datasets of both mass and precision practice. Third, we trained neural networks in DeepLabCut and refined the algorithms in python scripts.

### *2.6.1 Manual tracking using Tracker*

We used Tracker (version 5.1.5), a video analysis and modeling tool, to manually track the positions of the nose and paw of the mass practice video recordings. We calibrated the scale by utilizing the built-in calibration tool with the inner edges of the trough as a video feature with a known distance (9.5 cm). Next, the origin of the coordinate axes was positioned in the lower corner of the video frame, closest to the entrance of the reaching chamber. Furthermore, the axes were angled such that the x-axis was parallel to the trough. The experimenter then tracked the nose and paw by clicking on the tip of the nose and the 5$^{th}$ metacarpophalangeal joint, respectively. In addition, we labeled four landmarks of the reaching chamber that served as reference points to define the borders of the region of interest (ROI) (**Figure 3A**). The x- and y- components of position for all objects (i.e. reference points, nose, and paw) were obtained, as well as the velocity x-component of the paw ($v_x$), and the path length of the paw. All variables were measured in centimeters. Tracker calculated velocities from the position-time data using the finite difference method:

$$v_i = (x_{i+1} - x_{i-1})/(2 * dt) \quad (1).$$

Subscripts refer to frame numbers and dt is the time between frames in seconds. A positive value indicates a forward movement and negative values indicate a retracting movement. We used the reversal in the sign of $v_x$ to identify the end reach phase and the beginning of the retraction phase (see below). The path length was calculated using the Euclidean distance between two consecutive frames and subsequent summation of these segments.

2.6.1.1 Number of reaches

Exported Tracker data of 5 video samples (~10 s length) were used to develop algorithms to quantify the outcome measures of interest: number of reaches, time spent on task (s), and the total path length of the paw (cm). First, these videos were visually inspected for reaching events. The frame numbers of maximal paw extensions (last positive velocity) and 3 subsequent retraction frames were logged (n = 34 reaching events). Then, the paw velocity values corresponding to these frames were used to determine the initial threshold as follows. We calculated the average velocity across 3 retracting frames for each reaching-retracting event and these average numbers were sorted from fast to slow. The detection threshold was set at the value that included 95% of the average retracting velocities. Retraction events were only counted provided they met the boundary conditions. Namely, the x-component of the paw had to be between the vertical lines of the edges of the trough (i.e. mean of x-coordinates of reference points 1 and 3, and points 2 and 4, see **Figure 3B**). Thus, retraction events were counted if they met the following statements:

Retraction is True IF:

paw (x) < mean of reference points 1 and 3 (x) AND

paw (x) > mean of reference points 2 and 4 (x) AND

Sign ($v_{x,i}$) + Sign ($v_{x, i+1}$) = 0 AND

mean of 3 retracting frames velocities (x) ≤ threshold

2.6.1.2 Time on task

The variable time on task was assessed based on the position of the nose. The nose had to be located below the horizontal line defined by reference points 1 and 2 and between the vertical lines defined by reference points 1 and 3, and 2 and 4 (**Figure 3B**). These criteria are met when rats direct their attention to the pellets by sniffing them. The equation determined for each frame whether the subject was “on task” or “not on task” and total time on task was computed using the following statements:

Frame “on task” IF

nose (y) < mean of reference points 1 and 2 (y)

nose (x) < mean of reference points 1 and 3 (x)

nose (x) > mean of reference points 2 and 4 (x)

Frame “not on task” IF

nose (y) > mean of ref points 1 and 2 (y)

nose (x) > mean of ref points 1 and 3 (x)

nose (x) < mean of ref points 2 and 4 (x)

$$Time\ on\ task\ (s) = \frac{Number\ of\ \text{"on task"}\ frames}{fps} \quad (2)$$

<u>2.6.1.3 Path length</u>

Path length is the only variable not restricted to a region of interest. For this, the algorithm found the maximum value of the paw path length that was measured by Tracker.

$$Total\ path\ length = maximum\ value\ of\ path\ length\ (cm)$$

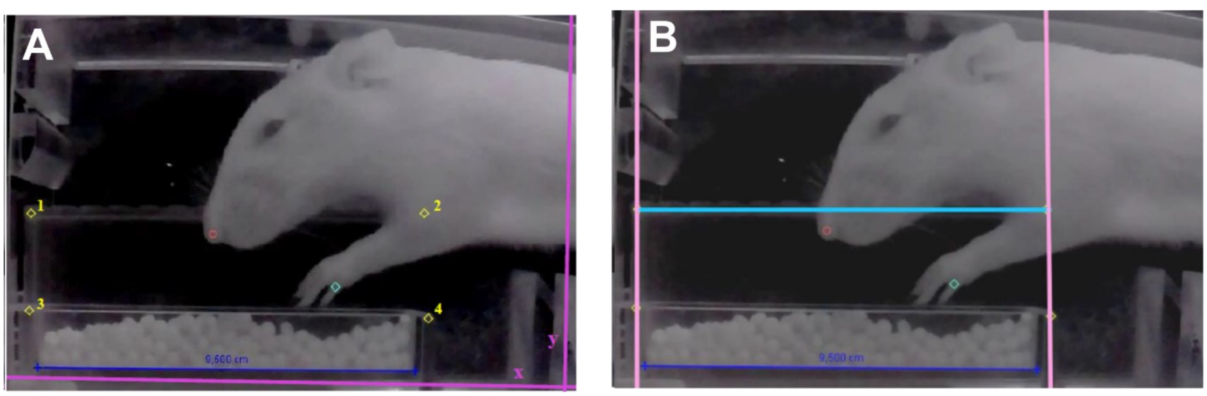


**Figure 3. Nose and paw position tracking in Tracker. A)** The nose and paw positions are indicated in red and green, respectively. The reference points 1, 2, 3, and 4 are indicated in yellow. **B)** Borders of the regions of interest cross through the reference points. The variable number of reaches is defined as when the paw retracted between the pink lines. The variable time on task is defined as when the nose was below the blue line and between the pink lines. The variable “sum of forelimb movements” is defined as the total distance traveled by the paw in the whole field of view in all frames of the video.

*2.6.2. Sampling error analysis*

Manual tracking is a lengthy process: the tracking of a 10-second video of the mass practice rehabilitation took around 30 minutes. For the amount of data collected in this experiment, 1 480 videos of 15 minutes length, the estimated tracking time would accumulate to 66 600 hours. This exemplifies the laborious nature of acquiring positional data through non-automated processes. Therefore, we investigated how analysis time could be reduced while preserving accuracy in our outcome measures of interest. To do this we made an ethogram of nine mass practice rehabilitation videos, classifying the rat's behavior as "active reaching", "resting", or "out of frame" for each second of the video. These data were then subjected to a sampling error analysis to determine the % deviation of the estimated value from the true value. With this analysis, we investigated how different methods of sampling the data (i.e. by varying time bins and frequency) would affect the amount of induced error.

Ahead of the analysis, the manual annotations from the ethogram regarding "active reaching" were compressed to a single entry per 10 seconds. Such that, if a rat was reaching 7 out of 10 seconds, the value 7 was used in the shortened entry list. See also **Figure S 1**. We then picked different time bins of interest, being 60, 30, 20, or 10 s. These bins represent the smallest segments of a subsample. We also determined different durations of the total sampled time, being 1, 3, or 5 minutes. Thus, to fulfill the total sampled time, several bins may need to be used, e.g. two bins of 30 seconds to have a total sample time of 1 minute. Next, we applied a sliding window technique, using the bins, to obtain multiple subsampled values per bin + sample time combination. These subsampled values were used to estimate the total value across the whole video:

$$Estimated\ value = subsampled\ value * \frac{video\ length}{total\ sample\ time}\ (3),$$

whereby video length and total sample time were inserted in minutes and subsampled value was the sum across all bins. Continuing with the example above, if the subsampled value across both bins was 27, then the estimated value = 27 * (15/1) = 405. Of the list of estimated values, obtained from the window positions, the minima and maxima were identified and used to define the range of percent error. The percent errors were calculated using this formula:

$$\%\ error = \frac{Estimated\ value - True\ value}{True\ value} * 100\%\ (4).$$

A positive percent error indicates that the data sample overestimates the true value of the original dataset, and a negative outcome implies an underestimation. Finally, the percent errors were averaged across animals.

### *2.6.3 Automated tracking in DeepLabCut*

Video recordings were processed by DeepLabCut (v. 2.1.8), using a Titan XP GPU. We modified and refined the Tracker algorithms to be able to process DLC output.

2.6.3.1 Dataset preparation

Prior to subjecting rat-reaching videos to DLC for automated pose estimation, we edited the videos using a Python script. The videos were 1) compressed from .h264 codec to MPJG and .AVI format; 2) spatially downscaled 3-fold, in height and width; 3) flipped horizontally if the rat was right-handed, enabling uniform pose estimation and post-processing; 4) divided into 1-minute chunks; 5) converted to grayscale.

2.6.3.2 Training

We trained two distinct, ResNet-50-based (He et al. 2016), neural networks for automated body part tracking in mass and precision practice videos, over 8 and 6 iterations respectively. Specifically, we used 2148 labeled frames taken from 57 mass practice videos (18 animals), whereby the nose, paw, and two inner edges of the pellet trough were labeled. For the precision practice network, we labeled 1397 frames taken from 37 videos (14 animals). This network was trained to track the nose, paw, edges of the pellet pedestal, and whisker pad. In both networks, 90% of the labeled frames were used for training (10% for testing) and the 'imgaug' image augmentation method was applied. The training frames were selected to capture behavioral and experimental variability, i.e. varying lighting conditions and poses.

2.6.3.3 Post-hoc processing

Video pixel coordinates were imported into Python. For each label's coordinates, DLC gives a confidence measure of the label location between 0 and 1, which is termed likelihood. Label positions with a likelihood <0.90 were replaced with 'Not a Number' values. Likewise, for unrealistically large jumps in paw positions between two frames. The displacement between paw positions was determined using the Euclidean distance.

As stated above, the previously established algorithm developed from Tracker output were refined. The origin of the coordinate axes was now positioned in the upper right corner of the video frame, closest to the entrance of the reaching chamber. The ROI in the mass practice videos was expanded 1.3 cm, to the bottom of the trough. Reach detection threshold was determined in a similar fashion as described above. However, reaches were now defined as an increasing distance away from the origin. Specifically, to determine the threshold, we calculated the change in Euclidean distance, relative to the origin, between two paw positions that were three frames apart:

$$\Delta d_{(i-(i-3))} = d_i - d_{i-3} \quad (5).$$

whereby $i$ is the frame showing maximum paw extension and $d$ is the Euclidean distance from the origin. The equation for time on task remained the same.

The ROI for reach detection in the precision reaching task extended bilaterally ~2.0 cm from the edge of the pedestal proximal to the reaching chamber, for the entire frame height. A second ROI was used to measure time on task. This extended ~4.0 cm laterally from the pedestal, into the reaching chamber, with a height extending from the top of the frame until the pedestal's peak cycle position. Reaching threshold was established by observing the change in paw position 2 frames prior to peak extension. Frames, whereby the paw was located in the first ROI or the nose or whisker pad in the second ROI, counted towards time on task.

**2.7 Statistical analysis**

Statistical analysis was performed using GraphPad Prism 9 (San Diego, CA, USA). Wilcoxon matched-pairs tests were used to compare human and algorithm analyzed data. Paired T-test were used to compare subsampled and fully analyzed data sets. An unpaired T-test was used to compare infarct volumes between mass and precision practice animals. To compare stroke sizes between the rehabilitation groups in each paradigm (4, vs. 8, vs. 12), we used a one-way ANOVA. Missing values in the collect video dataset were imputed in R version 4.1.1 (R coreteam, Vienna, Austria), using the MICE package (Multivariate Imputation via Chained Equations) (van Buuren and Groothuis-Oudshoorn 2011). Pre-stroke and day 1 rehabilitative training performances were compared using a two-way repeated measures ANOVA (time X group) with Šídák's post-hoc analysis. To analyze delivered training dose across rehabilitation, we used two-way repeated measures ANOVA (time X group) with Tukey's post-hoc correction for multiple comparisons. Grand mean and total dose after completion of the paradigms were compared between groups using a one-way ANOVA. Pearson correlation analysis was used to assess the relationships between reaches, time on task, path length, staircase test week 1, and change in staircase test (week 3-1), with pellets retrieved and between change in staircase test with infarct volume. A two-way repeated measures ANOVA (time X group) was used to compare staircase pellet retrieval (with contralateral forelimb) at the different time points and between groups. With Tukey's post-hoc analysis. A value of $p$ <0.05 was considered significant and data are presented as mean ± SD unless otherwise indicated.

## 3. RESULTS

### 3.1 Development of reach detection algorithms

We aimed to use the trajectory of the paw to create a formula for reach detection. Five mass practice video samples that were tracked in Tracker, containing 34 reaches in total, were used to determine the threshold to detect the retracting phase in the reach-to-eat sequence. We found that 95% of the retracting movements (of 3 frame duration) had an average velocity along the x-axis of -3.24 cm/s or greater (**Figure 4A**). Contrasting the algorithm for Tracker-derived data, the python script for DLC post-processing was designed to identify forward movements (i.e. the reach in the reach-to-eat sequence). We used the $5^{th}$ percentile of 88 reaching events derived from 10 mass practice video samples to determine the threshold above which movements should be counted as a reach, which was established at 0.325 cm. Similarly, the $5^{th}$ percentile boundary of 79 reaching events in precision practice determined the reaching threshold as 0.439 cm (**Figure 4B**). A visual representation of reaching events and detection thereof is depicted in **Figure 5**. The reaching movements of rats (above the threshold) typically last about 2-4 frames, which is equal to 0.07-0.14 s. That is, the Euclidean distance of the paw position in two frames (3 frames apart), was used to define our criterion for a reach. Typically, a reach is characterized by an initial increase of the Euclidean distance, indicating an increase in speed, followed by a decreasing distance, indicating deacceleration. Although the minimum threshold to detect a reach is 0.325 cm, the rat paw usually demonstrates a Euclidean displacement of 1 to 2 cm (**Figure 4B, left; Figure 5, left**). Events that are just above the threshold for a single frame tend to be falsely identified reaches by the algorithm, e.g. the second peak in **Figure 5**.

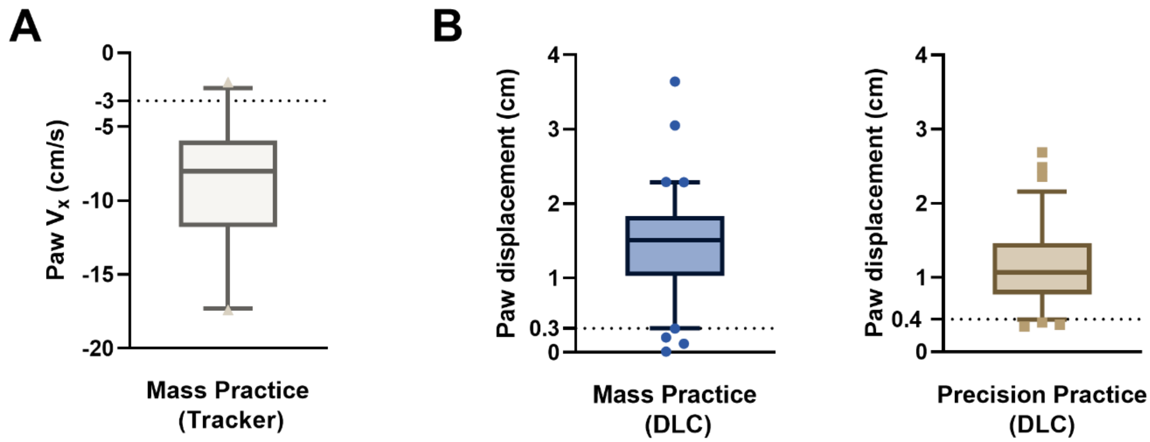


**Figure 4. Paw velocity and displacement following or preceding maximum limb extension. A)** Mass practice horizontal paw velocity measured over 3 frames (n=34). The retraction detection threshold was set at -3.24 cm/s. **B)** Mass practice paw displacement was measured over 3 consecutive video frames (n=88) and the reach threshold was determined at 0.325 cm. Precision practice paw displacement was measured over 2 consecutive frames (n=79) and the threshold was established at 0.438 cm. The horizontal line denotes the median value, boxes extend from the 25th to 75th percentile of the sample's distribution of values, whiskers extend to the 5th and 95th percentile. Dots denote values outside of the whisker range. The vertical dashed line indicates the reaching threshold, set at the 5th percentile.

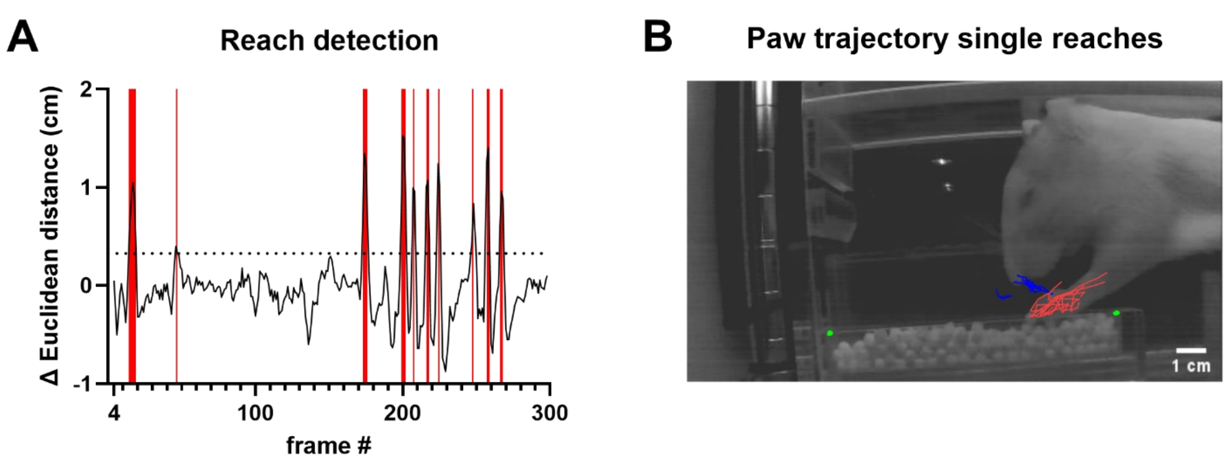


**Figure 5. Visualization of reach detection in mass practice task. A)** the change in Euclidean distance of the paw (relative to the origin) between current and third last frame, is graphed against the frame number. A reach was detected when the change in Euclidean distance was greater than the threshold of 0.325 cm, indicated by the vertical dashed line. The shaded red bars denote the beginning and end of a reach. The second peak in this graph is a false positive detection, recognizable by a tiny peak and short duration. **B)** shows paw trajectories during reaching movements in red, and the nose trajectory in blue. Green dots denote the inner edges of the trough.

### 3.2 Accuracy of reach detection by algorithms

To assess the accuracy of our algorithms in identifying reaching events we compared the number of reaches counted by human analysis with the computer output. For each modality, 30 video clips were evaluated by an experimenter and the associated algorithm. We compared the values by computing Wilcoxon matched-pairs tests. This test indicated that the median number of reaches recorded by Tracker was significantly lower than the median observed to human visual inspection, with a median of differences of -1.0 ($p = .0024$, 95% CI [-1.52, -0.28]). In contrast, the median of differences was +1.0 for DLC versus human analysis in the mass practice task, demonstrating that the DLC mass practice algorithm overestimates the number of reaches ($p = .0245$, 95% CI [0.04, 1.36]). The median number of reaches detected by the DLC precision practice algorithm was higher than the median recorded by the experimenter, however, the median of differences between these data sets was 0, showing no statistical difference ($p = .4103$, 95% CI [-0.46, 2.13]).

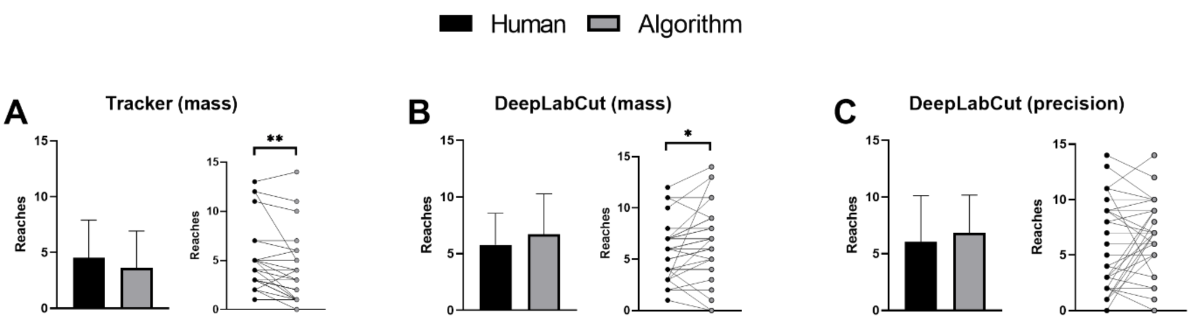


**Figure 6. Human and algorithm reach count comparison.** Three sets of video samples (n=30), duration 10 s, were scored by an experimenter and the corresponding algorithm. The left graph in each panel shows the average number of reaches (± SD) and the right graph depicts individual values. The lines between values indicate data pairs. **A)** The tracker-based algorithm for mass practice videos slightly underestimates the average number of reaches in a video sample. **B-C)** The DeepLabCut algorithm slightly overestimates the average reach count. Significance of $p < .05$ is denoted by a single asterisk (*), $p < .01$ is denoted by a double asterisk (**).

Another mathematical approach was used to further describe the performance of the algorithms. We adopted the metrics of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) that are often used in the medical field to determine the accuracy of tests with a binary outcome. These measures are based on the following parameters: true positive (i.e. “hit”), true negative (i.e. “correct rejection”), false positive (i.e. “false alarm”), false negative (i.e.

“miss”) (Loong 2003). The values of these parameters were derived from a selection of video samples of 10 s long (Tracker, n = 30; DLC (mass), n = 30; DLC (precision), n = 30); the results are summarized in **Table 1**. The mass practice algorithm for DLC output had the highest sensitivity, being able to detect 98.2% of the reaches that were also recorded through visual inspection. Mass practice (Tracker) and precision practice (DLC) algorithms had a sensitivity of 90.2% and 81.7%, respectively. The DLC algorithms were substantially better at correctly rejecting small movements that were not a reach. This is reflected by the high specificity rates of 93.4% (mass) and 86.9% (precision) versus 63.2% in the Tracker algorithm. The precision (i.e. PPV) at which events were labeled as “reach” by the algorithms was somewhat similar, ranging from 71.4 to 81.2%. Finally, all algorithms demonstrated an NPV of at least 84%, which is the rate of events labeled as “not a reach” to truly be not a reach. Sensitivity and specificity were the most important performance metrics to determine whether the algorithm was satisfactory for our experiments. The chief objective of the algorithm was to detect reaches, which is measured by sensitivity. Secondly, we wanted the algorithm to not mistake small jerky movements for reaches, indicated by specificity. Although sensitivity and specificity of 100% are ideal, in reality, this is impossible. We deemed a result of >80% acceptable for the purpose of this study, which allows us to relatively accurately and reliably estimate the number of reaches.

**Table 1.** Performance of binary classification of reaching events by algorithms.

| Parameter | Definition | Formula | Value Mass practice (Tracker) | Value Mass practice (DLC) | Value Precision practice (DLC) |
|---|---|---|---|---|---|
| **Sensitivity** | Percentage of reaches detected by algorithm also observed by human visual inspection | True Positive / (True Positive + False negative)*100% | 90.2% | 98.8% | 81.7 % |
| **Specificity** | Percentage of algorithm detected events as “not a reach” also (not) observed by human visual inspection | True Negative / (True Negative + False Positive)*100% | 63.2% | 94.5 % | 86.9 % |
| **Positive predictive value** | Likelihood an algorithm detected reach is truly a reach | True Positive / (True Positive + False Positive)*100% | 74.8% | 84.7% | 71.4 % |
| **Negative predictive value** | Likelihood an algorithm detected “not a reach” event is truly not a reach | True Negative / (True Negative + False Negative)*100% | 84.21% | 99.6% | 92.2 % |

### 3.3 Sampling error analysis

The sampling error analysis was initially conducted to determine how manual tracking time could be reduced while retaining accuracy in the estimated value of the outcome measure of interest. First, the distribution of rats' active reaching behavior throughout a mass practice training session was outlined, using 9 full-length (15-minute) videos. Most rats were actively reaching for pellets between the third and ninth minute of the training session **(Figure 7A**). After that, engagement declined and only ~10% of the animals continued to exhibit active reaching behavior. This was reflected in the cumulative percentage of total "active" frames (**Figure 7B**), showing that by the end of the ninth minute, approximately 90% of the rat's active reaching time was achieved. Based on these data, subsampling analysis used snippets (i.e. bins) of video that originated from multiple time points across the training sessions in order to capture both occasions of higher and lower activity. The subsample can vary both in bin size and the total amount of footage taken. Larger bin sizes and short total sample time resulted in a greater percent error (**Figure 7C**). With a bin size of 60 s, the negative and positive percent error was -100 to 255% at a total sample time of 60 s, and -30 to 38% at a total sample time of 300 s. The exact same total sample times (60 and 300 s), sampled with bins of 10 s, gave an error range of -38 to 57% and -5 to 5%, respectively. Thus, the sampling error analysis showed that frequently analyzing 10-second snippets across the video, such that it accumulates to a total sample length of 300 seconds, would result in the least amount of error. This is equal to sampling 10 s every half minute (**Figure 7D**). However, sampling at this rate is still laborious. To further assess the potency of subsampling we used the 10 s per minute sequence to compare extrapolated with original results in both a subset of DLC and visually analyzed videos.

The average estimated number of reaches using the sampling method was similar to the average number of reaches observed across full-length videos. No statistical test was performed on the Tracker-derived data due to the small sample size (n = 4). However, the mean of original data was 124.5 (SD = 141.5) and of the extrapolated data 142.5 (SD = 166.0), **Figure 8A**. Paired sampled t-tests were conducted to compare original data obtained from DLC and manual full video analysis with extrapolated data, obtained by applying the sampling method described above. The statistical analysis revealed no significant difference in the mean number of reaches observed by original (mean = 167.2, SD = 110.6) and extrapolated (mean = 164.5, SD = 106.5) data in the mass practice task (n = 18); $t(17) = 0.3579$, $p = .7249$, **Figure 8B**. This was also true for precision practice videos (n = 15), whereby the mean of original data was 240.5 (SD = 226.2) and extrapolated data 233.2 (SD = 218.7); t(14) = 1.650, $p$ = .1211, **Figure 8C**. Manually scoring the videos does not alter the outcome. Mass practice videos that were scored by hand (n = 9) did not demonstrate a significant difference between original (mean = 195.0, SD = 81.63), and extrapolated data (mean = 215.3, SD = 99.43), t(8) = 1.168, $p$ = .2764, **Figure 8D**. Likewise, the average number of reaches per training sessions manually observed in the precision reaching (n = 6) task was not different between original (mean = 335.3, SD = 178.7) and extrapolated (mean = 345.0, SD = 239.7) data; t(5) = 0.2576, $p$ = .8070, **figure 8E**.

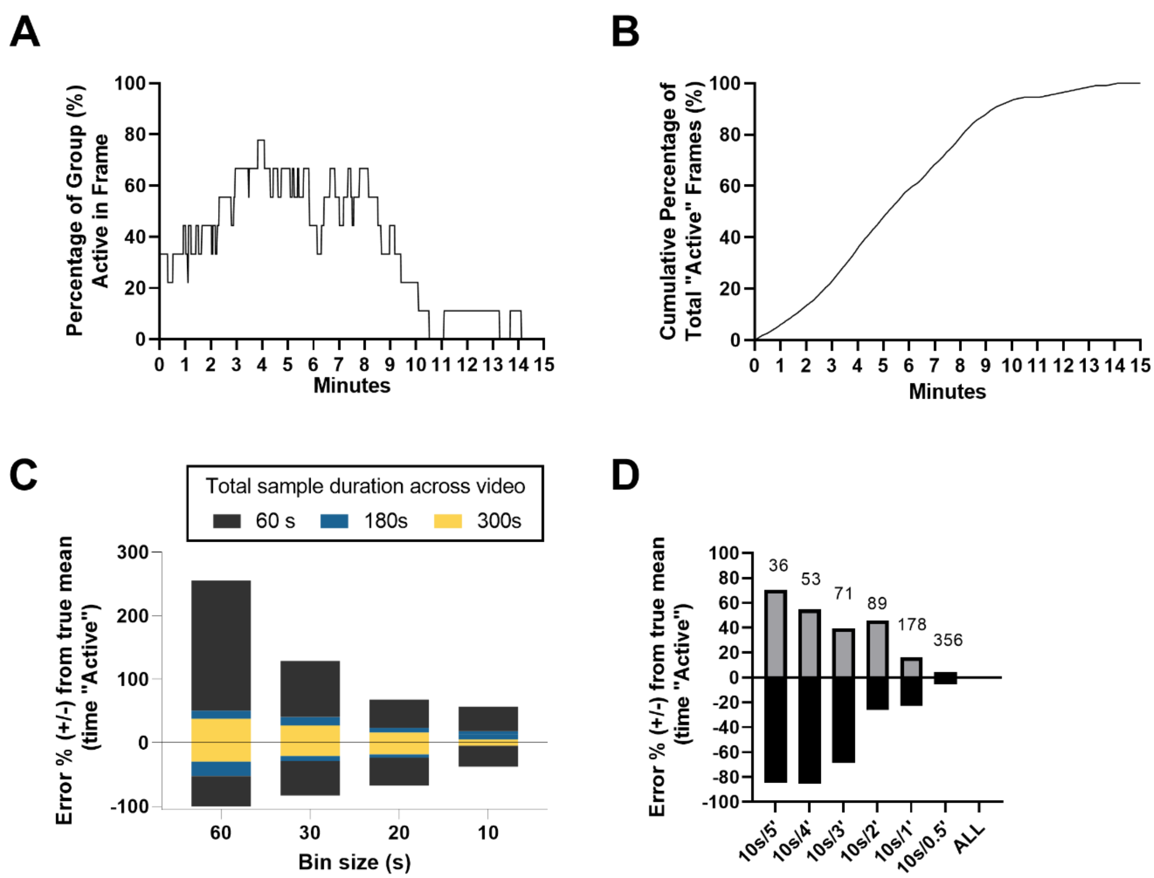


**Figure 7. Activity distribution and sampling error analysis**. Nine full-length mass practice videos were used to create an ethogram, classifying the behavior as "active reaching", "resting", or "out of frame". **A-B)** Shows the spread of rats' active reaching behavior throughout a rehabilitation session. After ~9 minutes, 90% of the rats' active reaching time was achieved. **C)** The error obtained from analyzing bins of video footage decreases when the total (summed) sample time is high, but bin size low. **D)** Potential error for different sampling spacings using 10-second bins. The numbers above the bars indicate minimal, manual, tracking time for a dataset of 76 videos.

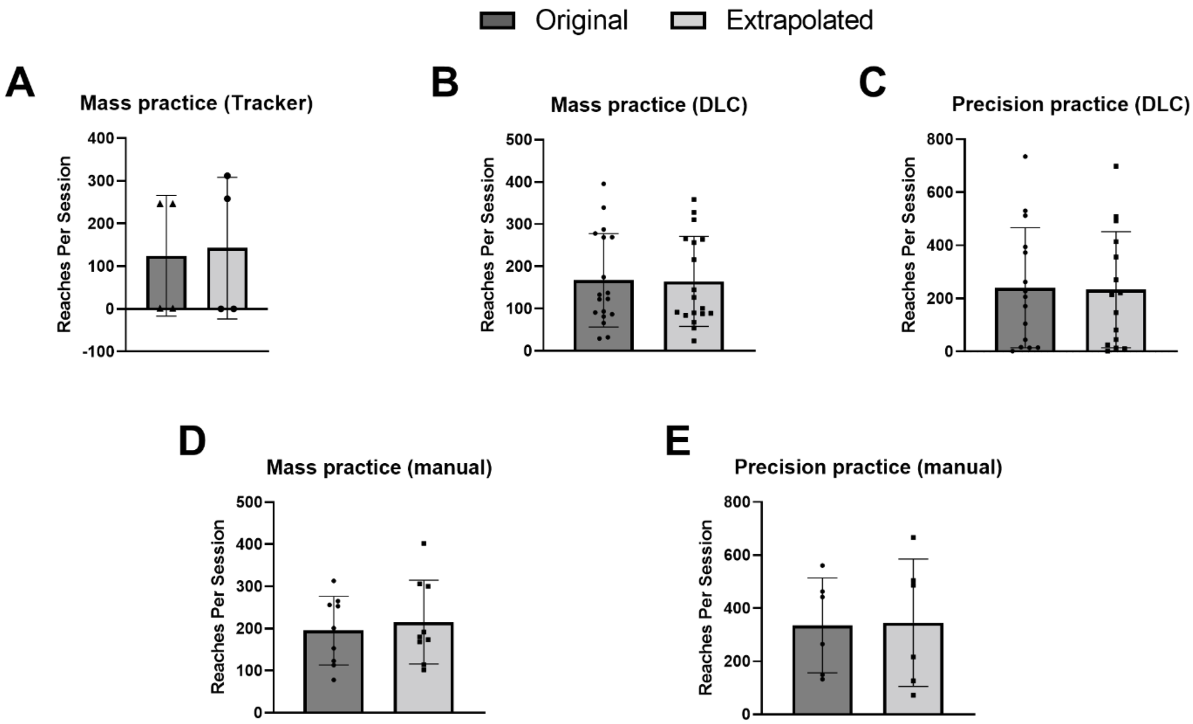


**Figure 8. Reach count in original and extrapolated data sets.** Extrapolated data were acquired by analyzing behavior for 10 seconds per 1 minute of video footage followed by multiplication by a factor of 6 to get a value that represents behavior during the entire session. The number of videos and rats used were as follows. **A)** Tracker, 4 videos from 1 rat. **B)** DLC – mass, 136 videos from 18 rats. **C)** DLC – precision, 119 videos from 15 rats. **D)** Manuel – mass, 9 videos from 9 rats. **E)** Manual – precision, 6 videos from 6 rats. Data are mean ± SD.

### 3.4 Photothrombosis induced unanticipated variable lesion volumes

Our photothrombotic stroke protocol generated variable infarct volumes, especially in the precision practice cohort. Mean lesion volumes in the mass practice and precision practice animals were 82.71 $mm^3$ ± 25.91 and 65.46 $mm^3$ ± 53.53, respectively. An unpaired t-test between the mass and precision practice animals revealed a non-significant difference in lesion volumes ($t(31) = 1.210$, $p = .2353$). Although the average lesion sizes were not different, the majority of precision practice animals (8/15) had a lesion volume < 54.0 $mm^3$ which was the smallest infarct size measured in mass practice animals (**Figure 9A**). Furthermore, the largest strokes in the precision practice cohort included some damage in the non-targeted hemisphere. In each paradigm, the rehabilitation groups were balanced based on the size of the infarct. Indeed, one-way ANOVA revealed that the mass practice rehabilitation groups were not significantly different from one another in lesion volume (F (2, 15) = 0.1339, *p* = .8757). Likewise, infarct size did not differ among precision practice training groups (F (2, 12) = 0.08343, *p* = .9205).

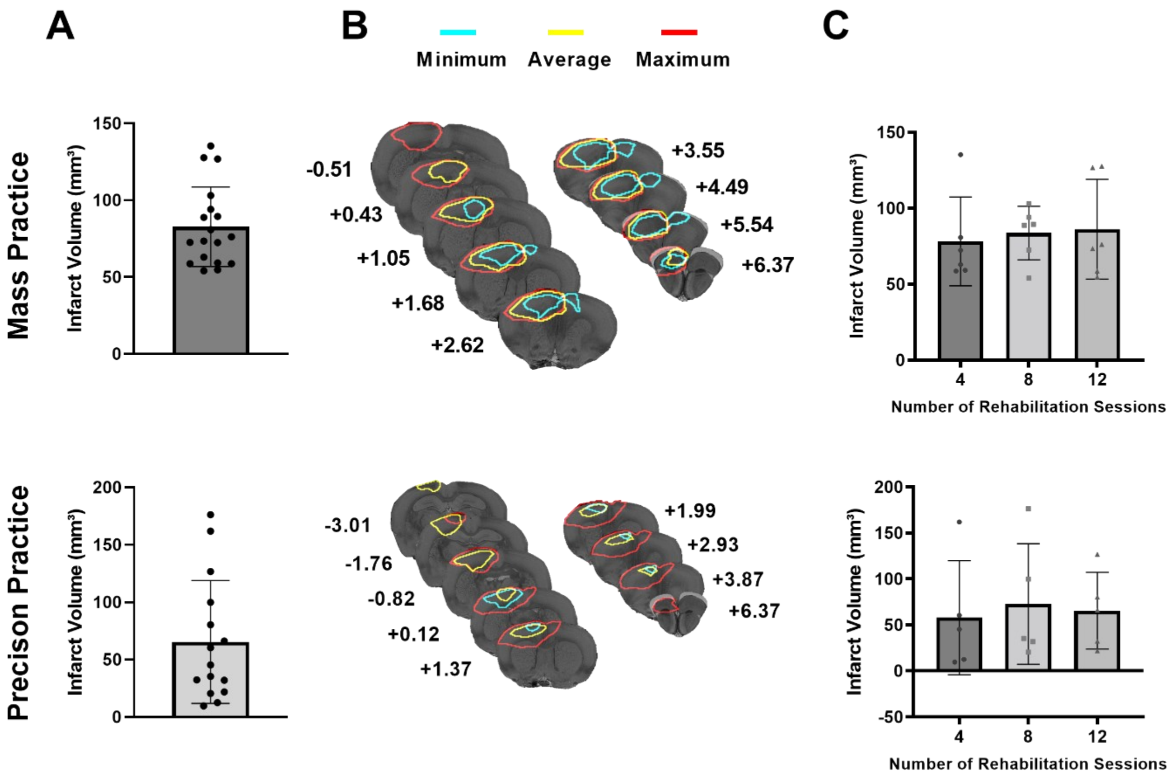


**Figure 9. Infarct volumes. A)** Animals had an average lesion volume of 82.71 ±25.91 $mm^3$ and 65.46 ± 53.53 $mm^3$ in the mass practice (n=18) and precision practice (n=15) experiments, respectively. **B)** Coronal images of the largest (red), mean (yellow), and smallest (cyan) lesion volumes after photothrombotic stroke. Illustrations show the extent of injury for antero-posterior coordinates (mm) **C**) Average lesion volumes per training group. Data are mean ± SD.

### 3.5 Reaching performance is impaired after stroke

*3.5.1 Mass practice*

Rats were trained on the rehabilitation task for several days prior to the stroke injury. Seven days following the surgery, the rehabilitation paradigms commenced. A two-way ANOVA was performed to analyze the effect of time (pre-stroke vs. rehabilitation day 1) and number of daily training sessions on grams consumed in the mass practice paradigm (**Figure 10**). The statistical test revealed there was a significant interaction effect between time and number of training sessions ($F (2, 15) = 4.407$, $p = .0312$). Simple main effects analysis showed that time did have a statistically significant effect on grams consumed ($p < .0001$) as did the number of training sessions ($p = .01$). Šídák's multiple comparisons test showed that all rehabilitation groups demonstrated a significant decline in grams consumed after the stroke (4 sessions: -1.517 g, $p = .0235$; 8 sessions: - 3.479 g, $p < .0001$; 12 sessions: -3.092 g, $p < .0001$). The pre-stroke training sessions of this task were not filmed, therefore comparisons on other metrics of dose were not possible for this time point.

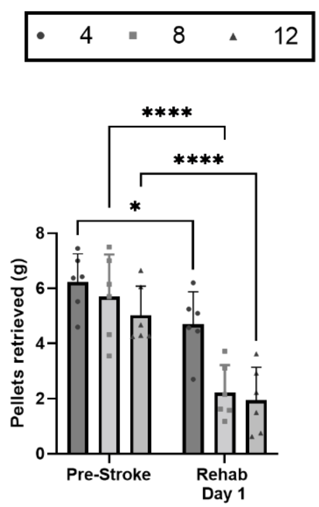


**Figure 10. Pellets consumed in the mass practice pre-stroke and at rehabilitation day 1.** All groups showed a significant decline following the stroke, at the first day of rehabilitation. The data shows the average of 4 pre-stroke training trials with the trough at the low height, and 4 training sessions of the first rehabilitation day that all groups performed simultaneously. Data are mean ± SD. Significance levels are depicted by asterisk: * $p < .05$; **** $p < .0001$.

*3.5.2 Precision practice*

The pre-stroke precision reaching training sessions, however, were filmed, and thus we compared all our metrics of dose before and after stroke. Before the stroke, the rats in the 4 sessions/day group consumed 1.614 g of pellets, rats with 8 sessions ate 1.423 g, and rats with 12 sessions retrieved 1.254 g. A two-way ANOVA revealed that there was no interaction effect between time and rehabilitation group, or a main effect of group. However, time has a significant effect on pellets consumed ($F(1, 12) = 40.18$, $p < .0001$). All rehabilitation groups demonstrated a significant decline after the stroke (**Figure 11A**; 4 sessions: 0.9760 g, $p = .0145$; 8 sessions: 0.6870 g, $p = .0055$; 12 sessions: 0.5940 g, $p = .0117$, Šídák's multiple comparisons test). Similarly, time has a significant effect on the number of reaches ($F(1,12) = 29,96$, $p = .0001$). Following stroke, the average number of reaches in a training session decreased by 442, 395, and 448 for rats with 4, 8, and 12 sessions respectively (**Figure 11B**; $p = .0203$; $p = .0385$; $p = .0187$; Šídák's multiple comparisons test). A two-way ANOVA also confirmed the effect of time on, time on task ($F(1,12) = 40.87$, $p < .0001$). Rats spent an average of 4 to 5 minutes less on their task after the stroke, leaving them engaged for only 2 to 4 minutes on the first day of rehabilitation (**Figure 11C**; 4 sessions, $p = .0047$; 8 sessions, $p = .0096$; 12 sessions, $p = .0175$; Šídák's multiple comparisons test). Finally, the reductions in pellets retrieved, reaches, and time on task is also reflected in the total amount of forelimb use. The average path length of the paw in a training session decreased significantly at rehabilitation day 1 compared to pre-stroke training ($F(1,12)$, $p = .0003$, two-way ANOVA). However, this diminishment was not significant for animals with 4 sessions per day ($p = .0590$). Rats with 8 and 12 sessions shortened their total path length by at least half of their baseline length (**Figure 11D**; $p = .0339$ and $p = .0353$ respectively, Šídák's multiple comparisons test). Thus, overall reaching performance in the rehabilitation tasks was significantly altered post-stroke.

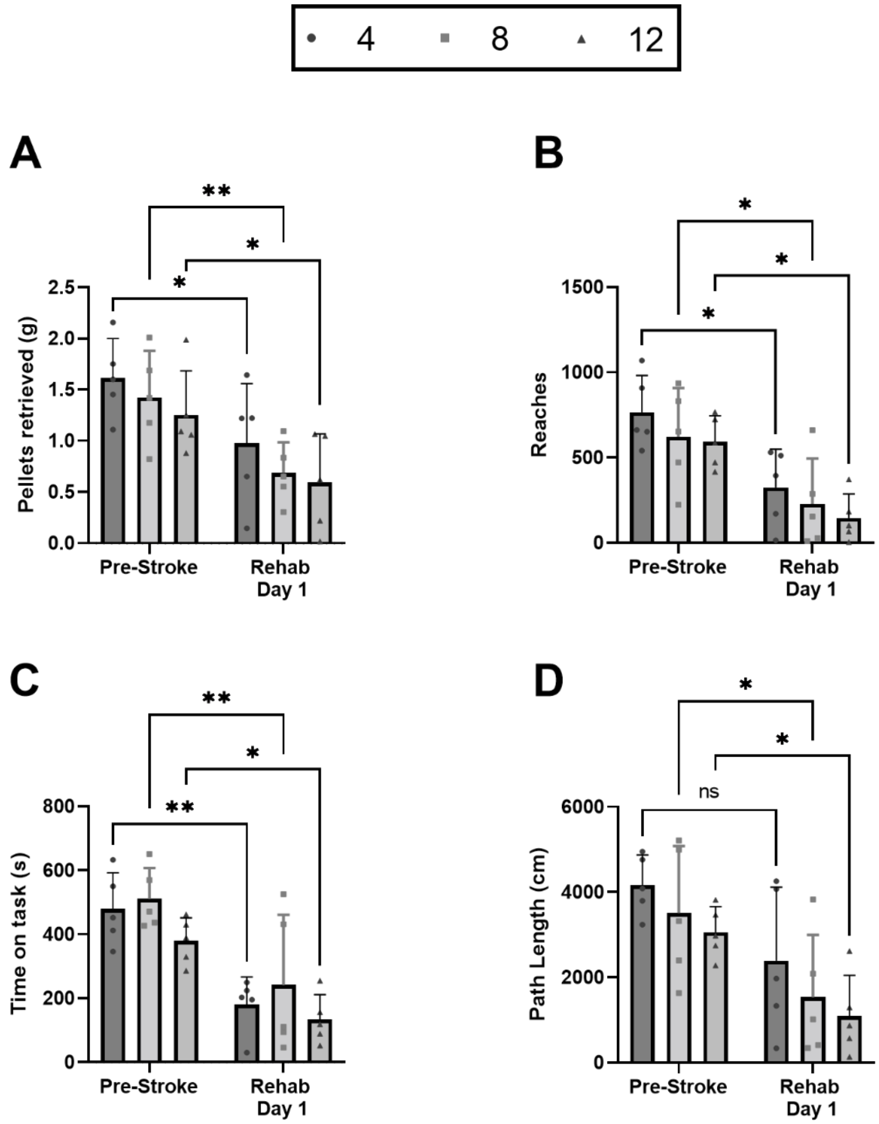


**Figure 11. Performance in the precision reaching task pre-stroke, and at rehabilitation day 1.** Data from rehabilitation days represents 4 training sessions, such that it matches baseline training that consisted of 4 sessions per day. Data are mean ± SD. Post-hoc Šídák's multiple comparisons effects are denoted by asterisks: * p <0. 5, ** P <0.01.

### 3.6 Training frequency does not impact accumulated rehabilitation dose but influences dose in independent sessions

*3.6.1 Mass practice – accumulated dose*

Rehabilitative forelimb training was given for 10 days whereby groups of rats received either 4, 8, or 12 training sessions per day. To assess whether daily accumulated reaching performance differed among groups and whether reaching practice changed throughout the rehabilitation period, we performed two-way repeated measures ANOVAs with Tukey's correction for multiple comparisons. The statistical test revealed that, in the mass practice rehabilitation paradigm, there was an interaction effect between the effects of time and rehabilitation group on the amount of pellets consumed ($F(18, 135) = 1.812$, $p = .0296$). The total amount of pellets consumed at the end of each rehabilitation day remained stable, with the minor difference that in the group with 4 daily training sessions, grams retrieved was greater at day 3 vs. day 9 ($p = .0176$) (**Figure 12A**). There was no interaction effect between time and effect of rehabilitation group on the daily summed value of the other metrics of dose (number of reaches, time on task, and path length). However, there was a main effect of time on time on task. The group with 4 sessions showed a dip in their task practice around day 7 to 9. Post-hoc pairwise comparisons confirmed reduced time on task in the 4 sessions group between day 1 and 8 ($p = .0013$), day 1 and 9 ($p = .0434$), and day 3 and 8 ($p = .0120$), (**Figure 12C**). Furthermore, rats with 8 sessions demonstrated a decrease in time on task between day 2 and 3 (**Figure 12C**; $p = .0463$).

The fact that there was no effect of rehabilitation group on the outcome measures across the rehabilitation days suggests that the groups' total performances in the paradigm were not different. This was indeed confirmed by one-way ANOVAs. Animals with 4 sessions retrieved 159.8 g ± 19.9 of pellets, animals with 8 sessions 181.4 g ± 34.9, and animals with 12 sessions 180.3 g ± 38.4 (**Figure 12A**; $F(2,15) = 0.8651$, $p = .4410$). Likewise, a non-significant effect of group was observed in the total number of reaches across the rehabilitation period (**Figure 12B**; $F(2,15) = 0.01844$, $p = .9818$). Rats performed a total of 8992 ± 1689 reaches with 4 sessions, 8914 ± 3583 with 8 sessions, and 8674 ± 3332 with 12 sessions per day. Total time on task was also not affected by rehabilitation dose (**Figure 12C**; $F(2,15) = 0.1755$, $p = .8407$). The group with 4 sessions had an accumulated time on task of 6786 s ± 2049, the group with 8 sessions 6684 s ± 2511, and rats with 12 sessions 7392 s ± 2126. Among the rehabilitation groups, the total path length of the paw throughout the rehabilitation paradigm was also similar (**Figure 12D**; $F(2,15) = 1.094$, $p = .3601$). Total path lengths were 50,541 cm ± 18,693, 53,869 cm ± 26,961 and 66,645 cm ± 10,612, for 4, 8, and 12 daily training sessions respectively. These data show that independent of the number of training sessions, rats received a similar dose after 10 days of training in the mass practice paradigm.

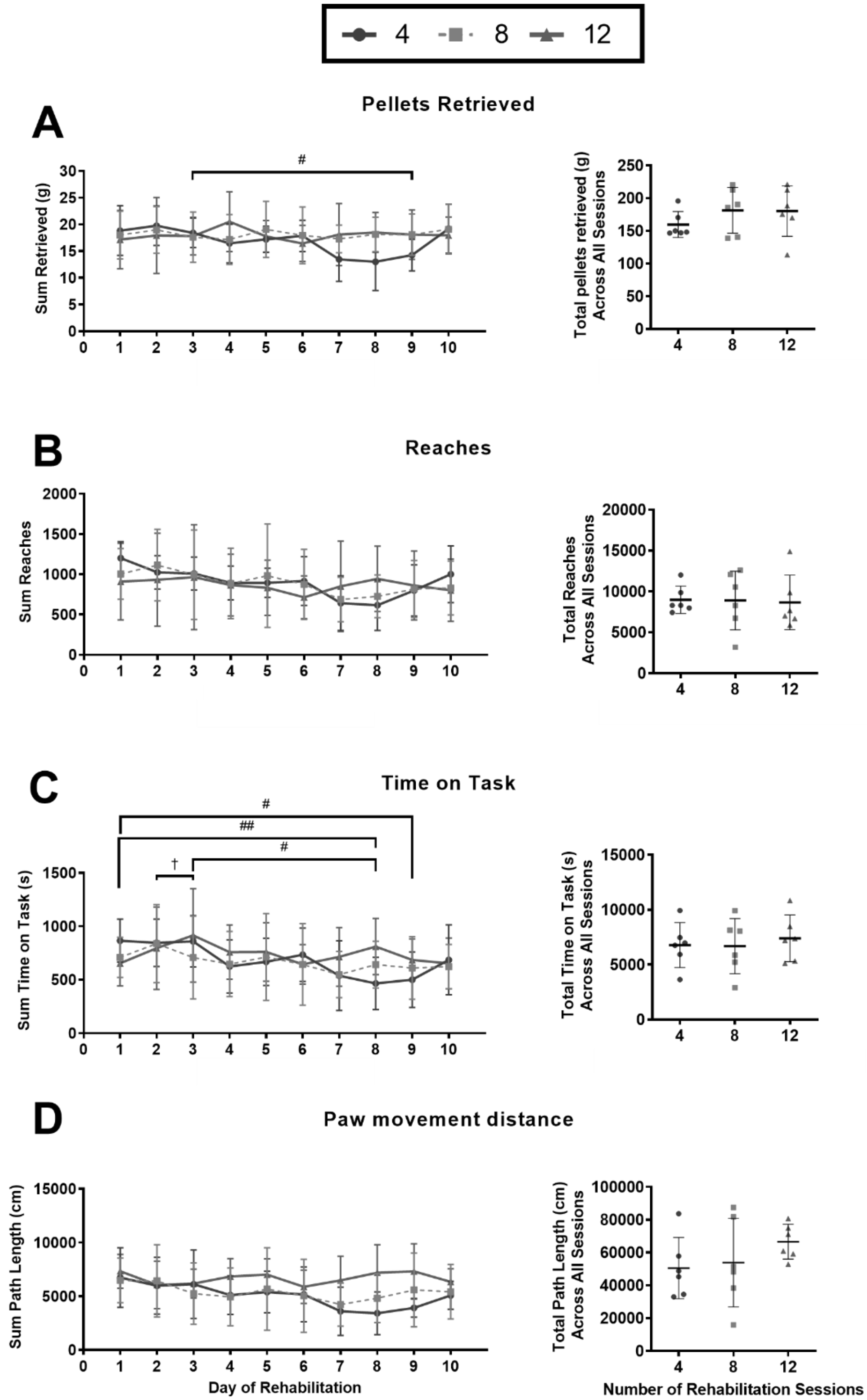


**Figure 12. Summed values of mass practice reaching rehabilitation.** Animals received 4, 8, or 12 sessions per day of reaching rehabilitation for 10 days. The left graphs show the daily summed value of the outcome measure across rehabilitation days. The right graph shows the total value across rehabilitation. # $p < .05$, ## $p < .01$ indicates post-hoc effect in group with 4 sessions, † indicates post-hoc effect in group with 8 sessions. Data are mean ± SD.

*3.6.2 Mass practice – session mean*

Although the accumulated training dose does not vary overall, we found that the rehabilitative training dose in single sessions varies among the groups. In contrast to the daily summed values described above, there was a significant interaction between time and rehabilitation group on the mean session value for each outcome measure (**Figure 13A-D**; grams retrieved: $F (18, 135) = 2.974$, $p = .0002$; reaches: $F (18, 135) = 2.075$, $p = .0098$; time on task: $(F18, 135) = 2.856$, $p = .0003$; path length: $F (18, 135) = 2.935$, $p = .0002$; two-way repeated measures ANOVA).

Over the course of the rehabilitation period, rats with 4 sessions per day performed fewer reaches per session, spent less time on task, and demonstrated a decrease in path length compared to rats with 8 or 12 sessions per day. However, there was an upward deflection during the last two days (**Figure 13**; 4, 8, 12 respectively: $F (2.200, 32.99) = 4.467$, $p = .0166$; $F (1.805, 27.08) = 5.194$, $p = .0145$; $F (1.446, 21.70) = 4.823$, $p = .0275$; RM two-way ANOVA). Furthermore, independent of the rehabilitation day, the group with 4 sessions always had a higher average value on the outcome measures compared to the groups with 8 and 12 daily training sessions. This main effect of rehabilitation group was confirmed by a two-way repeated measures ANOVA (grams retrieved: $F (2, 15) = 51.67$, $p < .0001$; reaches: $F (2, 15) = 24.77$, $p < .0001$; time on task: $F (2, 15) = 14.96$, $p = .0003$; path length: $F (2, 15) = 7.625$, $p = .0052$). Because the mean values are a fraction of the summed values, post-hoc tests showed discrepancies between daily averages for the same groups at the same time points as described above (grams retrieved: 4 sessions group, day 3 vs. day 9, $p = .0174$; reaches: 4 sessions group, day 1 vs. day 3, $p = .0035$, day 1 vs. day 8, $p =0.0445$, day 8 vs. day 10, $p = .0228$; time on task: 4 sessions group, day 1 vs. day 8, $p = .0013$, day 1 vs. day 9, $p = .0434$, day 3 vs. day 8, $p = .0120$; 8 sessions group, day 2 vs. day 3, $p = .0463$; path length: 4 sessions group, day 1 vs. 7 $p = .0288$, day 2 vs. day 6 $p = .0082$; Tukey's test).

Comparing the average amount of retrieved pellets per session among groups (**Figure 13A**), we found that the rehabilitation group indeed has a significant effect ($F (2, 15) = 51.85$, $p < .0001$, one-way ANOVA). Across all rehabilitation sessions, animals with 4 sessions per day consumed an average of 4.193 g ± 0.62 per session which was significantly more than animals with 8 (2.268 g ± 0.43; $p < .001$, Tukey's test), and 12 sessions (1.502 g ± 0.32; $p < .001$, Tukey's test). The difference in pellets retrieved between groups with 8 and 12 sessions was also significant ($p = .0330$, Tukey's test). During the training sessions, rats with 4, 8, or 12 bouts performed an average of 225 ± 42.2, 111 ± 44.8, and 72 ± 27.8 reaching movements, respectively (**Figure 13B**), showing that rehabilitation dose affects the number of reaches per sessions (one-way ANOVA $F(2, 15) = 24.77$, $p < .0001$). Like grams retrieved, animals with 4 sessions reached significantly more per session than animals with 8 ($p = .0004$, Tukey's test) and 12 sessions ($p < .0001$, Tukey's test). However, group differences between 8 and 12 sessions were not significant ($p = .2238$, Tukey's test). The animals with 4 sessions also spent the most time on task per session, namely 170 s ± 51.2, which is equal to ~18.9% of the sessions time (**Figure 13C**). This was more than animals with 8 sessions, who spent 84 s ± 31.4 (~9.3%) on task per session ($p = .0024$, Tukey's test) and animals with 12 sessions with an average of 62 s ± 17.7 (~6.9%) on task per session ($p = .0003$, Tukey's test). On average animals with 4 sessions per day moved their paw 1264 cm ± 467.3 per session, exceeding the path length of the group with 8 (673.4 cm ± 337.0, $p = .0214$, Tukey's test) and 12 sessions (555.4 cm ± 88.4, $p = .0064$) (**Figure 13D**). In summary, rats with 4 sessions per day retrieved more pellets, performed more reaches, spent more time on task, and moved their paw more, per training session. Rats with 8 and 12 sessions showed equivalent mean values for most dose metrics, except, animals with 8 sessions retrieved more pellets than animals with 12.

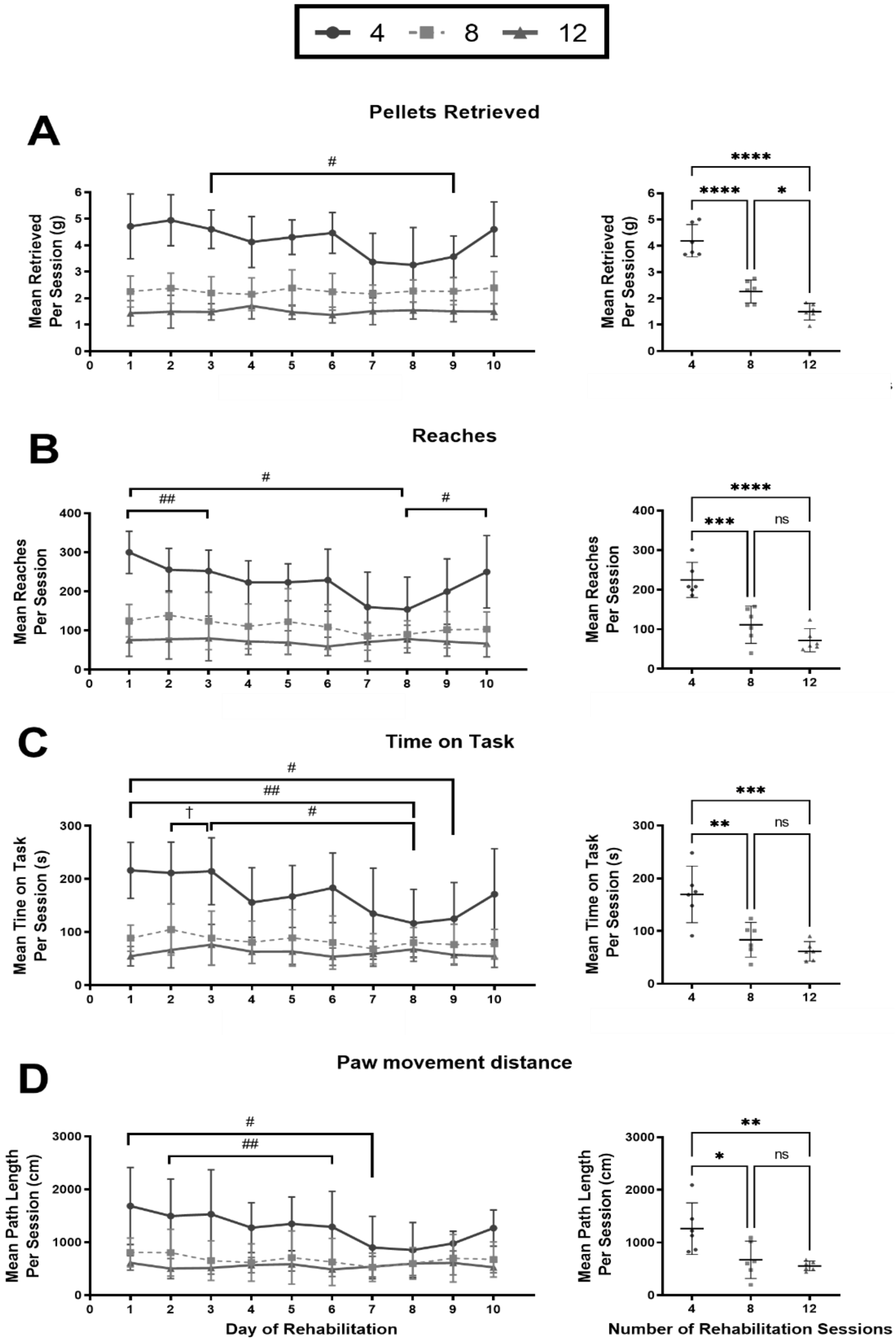


**Figure 13. Mean values per session of mass practice reaching rehabilitation.** Animals received 4, 8, or 12 sessions per day of reaching rehabilitation for 10 days. The left graphs show the mean value of the outcome measure per session, across rehabilitation days. The right graphs show the grand session mean across rehabilitation. # $p < .05$, ## $p < .01$ indicates post-hoc effect in group with 4 sessions, $p < .05$ † indicates post-hoc effect in group with 8 sessions, * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$, ns = non-significant. Data are mean ± SD.

*3.6.3 Precision practice – accumulated dose*

Throughout 10 days of rehabilitative training, all groups in the precision reaching paradigm achieved amounts of daily total of dose that were statistically not different (**Figure 14A-D**; Two-way repeated measures ANOVA). Group differences in the accumulated pellets retrieved across all training sessions were not significant (F (2,12) = 0.6537, *p* = .5377, one-way ANOVA, **Figure 14A**). Rats with 8 sessions retrieved the most pellets, 78.31 g ± 39.5, followed by the 12 sessions (61.68 g ± 31.1) and 4 sessions (54.48 g ± 30.0) groups. This was achieved by performing a total of 21,125 ± 11,458 (8 sessions), 14,648 ± 9,419 (12 sessions), and 12,688 ± 7279 (4 sessions) reaching movements, **Figure 14B**. Although animals with 8 sessions (21,944 s ± 10,823) spent more than twice as much time on task compared with the 4 sessions group (8,229 s ± 4,427), no group differences were revealed by statistical analysis (12 sessions (16,141 s ± 10,759); F (2,12) = 2.816, *p* = .0994, one-way ANOVA, **Figure 14C**). Finally, the results of the previous metrics are also reflected by the total paw use. Summed path lengths were: 127,706 cm ± 56,809 (8 sessions), 99,358 ± 56,079 (12 sessions), 86,520 ± 47,356 (4 sessions), F (2,12) = 1.098, *p* = .3601, **Figure 14D**. Although no group differences were revealed, these data suggest that rats with 8 training sessions per day may achieve the most forelimb use in the precision reaching paradigm. However, performance was variable, as some rats barely performed any reaches, resulting in large standard deviations.

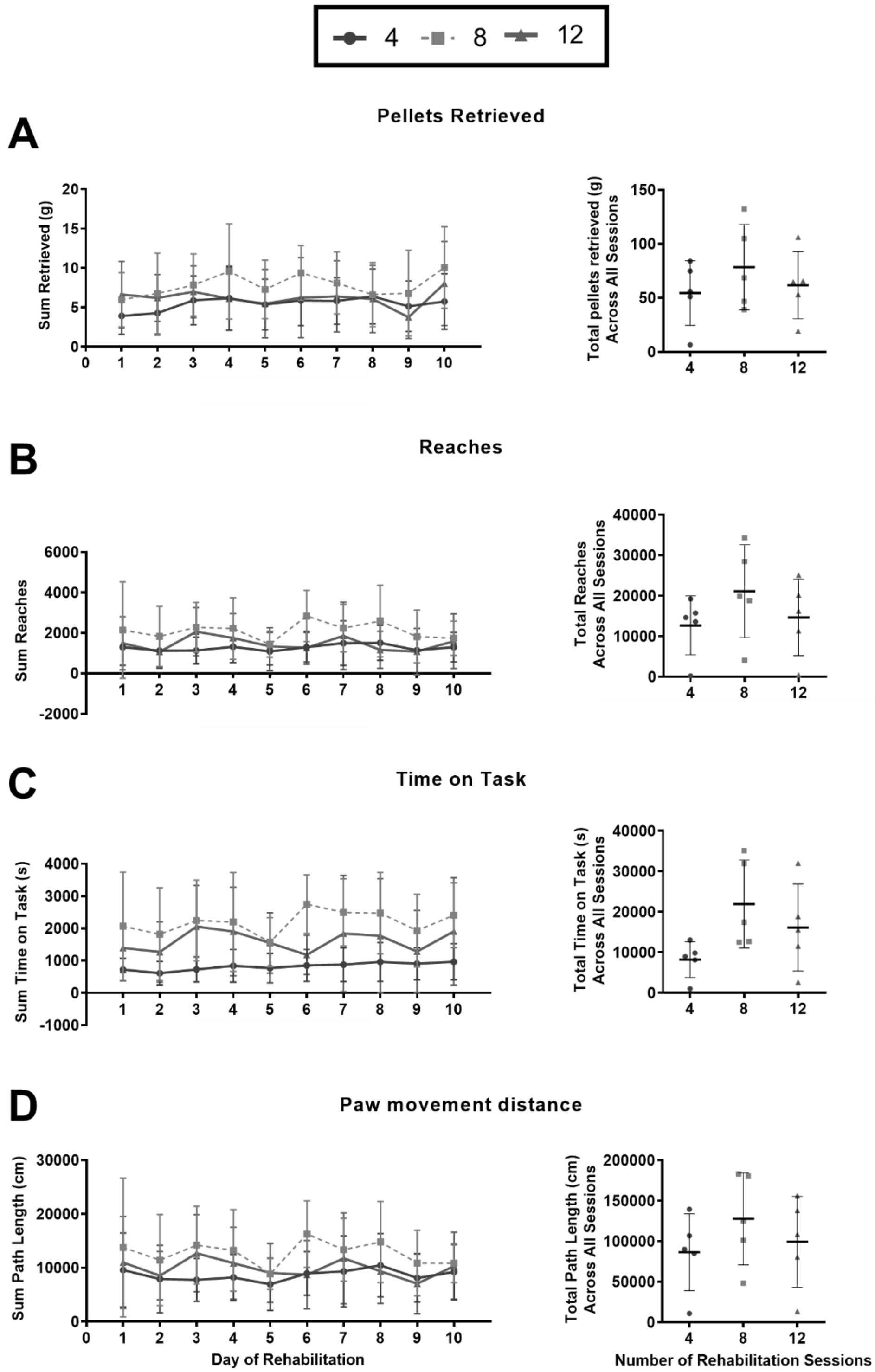


**Figure 14. Summed values of precision practice reaching rehabilitation.** Animals received 4, 8, or 12 sessions per day of reaching rehabilitation for 10 days. The left graph in each panel shows daily summed value of the relevant outcome measure across rehabilitation days. The right graph shows the total value across rehabilitation. Data are mean ± SD.

*3.6.4 Precision practice – session mean*

Across rehabilitation paradigms there were no interaction or main effects of time and rehabilitation group on the session mean value of the dose metrics (two-way ANOVA), indicating consistent performance. However, a small increase in mean pellets retrieved in the group with 4 sessions from day 1 to 3 of rehabilitation (**Figure 15A**) suggests that the animals improved their reaching accuracy given that other metrics remained unchanged. Different from the mass practice paradigm, the grand mean values of the dose outcome measures were not statistically significant among the rehabilitation groups (one-way ANOVA), likely due to a greater amount of variability in task performance. However, rats with 4 sessions per day retrieved the most pellets per session in this paradigm as well (1.362 g ± 0.75), followed by the groups with 8 sessions (0.979 ± 0.49) and 12 sessions (0.518 ± 0.25). This required 317 ± 1820 reaches with 4 daily sessions, 264 ± 143.2 with 8 sessions, and 122 ± 78.5 with 12 sessions per day (**Figure 15B**). Interestingly, the rats with 8 sessions spent the most time engaged with the task (274 s ± 135.3, (30.4%)). Rats doing 4 sessions spent 206 s ± 110.7 (22.9%) on task, while the group with 12 sessions spent 135 s ± 89.7 (15%) (**Figure 15C**). The average path length of the paw per training bout for rats doing 4, 8, or 12 sessions per day was 2,163 cm ± 1,184, 1,596.0 cm ± 710.1, 828 cm ± 467.3, respectively (**Figure 15D**). The difference between 4 and 12 sessions was nearly significant with a *p*-value of 0.0660. Thus, animals that received a low training frequency demonstrated higher mean activity in a single session. However, as their total number of sessions was also lower, the accumulated values of the dose metrics were either equal or lower than the groups with medium (8) or high (12) training frequency.

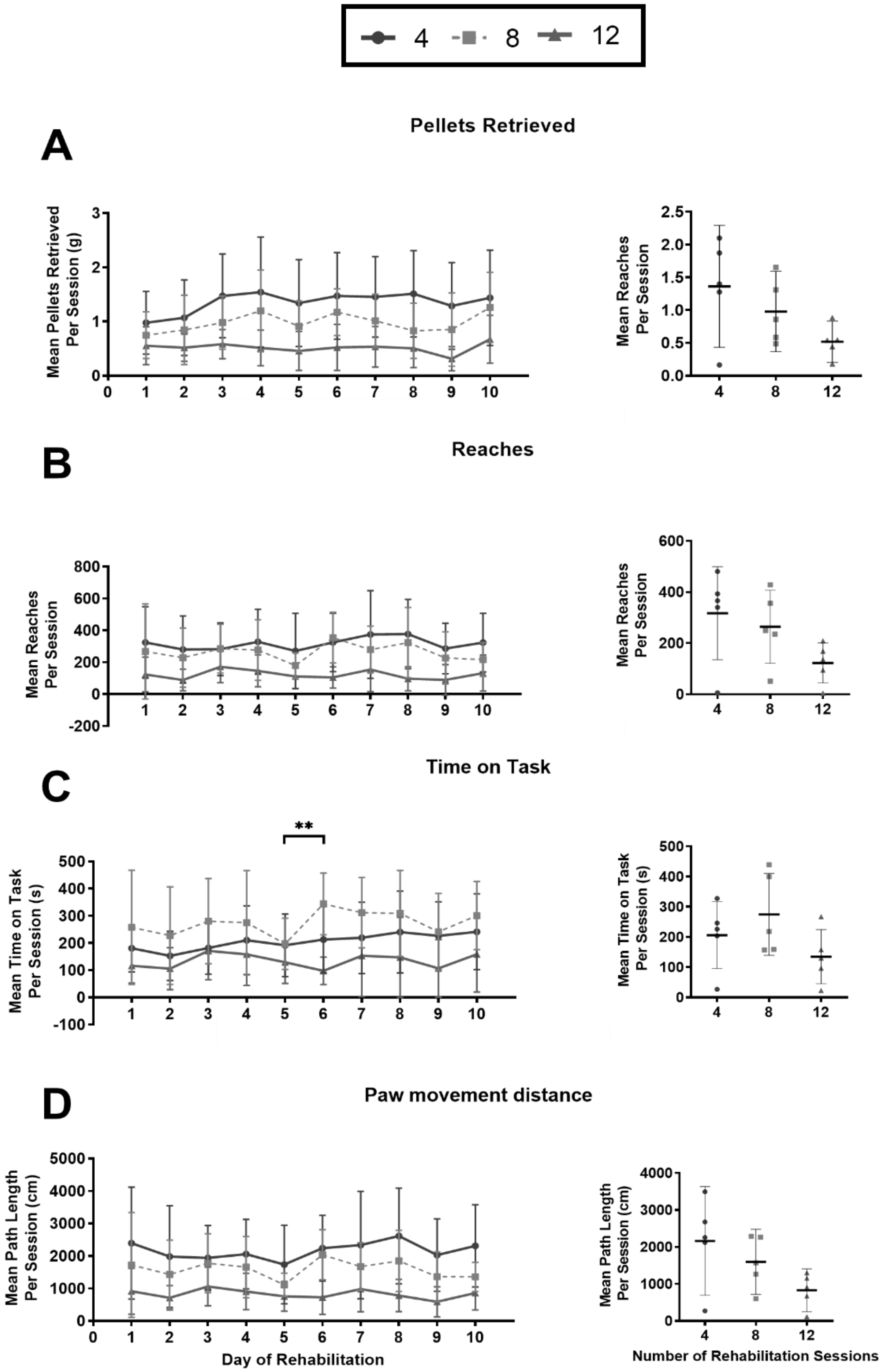


**Figure 15. Mean values per session of precision practice reaching rehabilitation.** Animals received 4, 8, or 12 sessions per day of reaching rehabilitation for 10 days (n = 5). The left graph in each panel shows mean value of the relevant outcome measure per session, across rehabilitation days. The right graph shows the grand session mean across rehabilitation. ** $p < .01$ indicates post-hoc effect in group with 8 sessions. Data are mean ± SD.

### 3.7 Differential correlations between video-derived dose metrics and pellet consumption in mass and precision practice

We assessed the relationship between the metrics derived from video analysis: number of reaches, time on task (s), paw path length (cm) with pellets retrieved (g), in both mass and precision practice data. Statistically significant correlations were revealed between number of reaches and pellets retrieved (**Figure 16A**, mass practice, $r(18) = 0.64$, $p = .0044$; precision practice, $r(15) = 0.80$), $p = .0004$). Time on task was not correlated with pellets retrieved in the mass practice task (**Figure 16B**, $r(18) = 0.3339$, $p = .1757$), but was in precision practice rehabilitation ($r(15) = 0.70$, $p = .0038$). The amount of forelimb use during mass practice was not correlated with pellets retrieved (**Figure 16C**, $r(18) = 0.1064$, $p = .6743$). In contrast, there was a strong correlation between the paw's path length and pellets retrieved in precision reaching ($r(15) = 0.8547$, $p < .0001$).

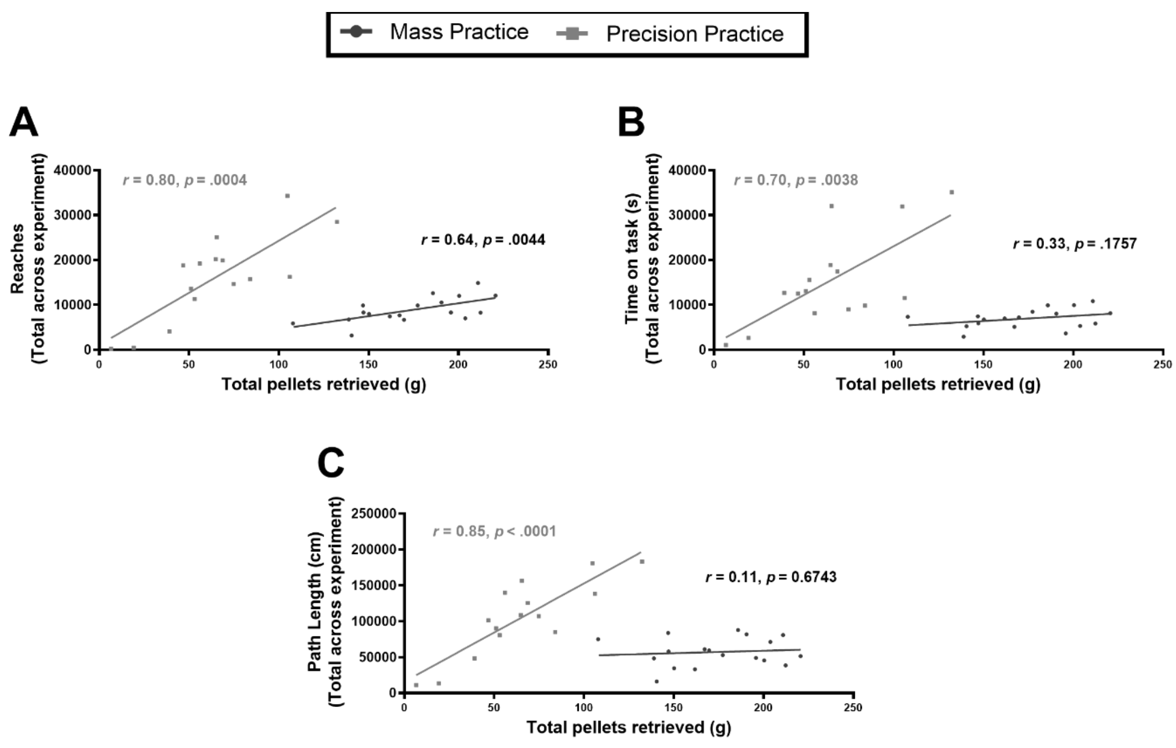


**Figure 16. Correlation analysis between video-derived metrics of dose and pellet consumption**. **A)** Reaches correlated with the pellets retrieved (g) in both the mass ($r = 0.64$, $p = .0044$) and precision practice task ($r = 0.80$, $p = .0004$). **B)** Time on task (s) did not correlate with pellets retrieved (g) in mass practice ($r = 0.33$, $p = .1757$), but was positive related in precision practice (r = 0.70, $p = .0038$). **C)** Path length (cm) was not correlated with pellets retrieved (g) in the mass practice task ($r = 0.11$, $p = .6743$) but was positively correlated in precision practice ($r = 0.85$, <0.0001). Sample size was (n = 18) mass practice animals and (n = 15) precision practice animals. Dashed lines indicate the 95% confidence intervals of the regressions.

## 3.8 Rehabilitation did not improve performance on Montoya staircase task

*3.8.1 Skilled reaching ability across time*

Skilled reaching ability was assessed with the Montoya staircase task at three time points in this study: before stroke induction (baseline), before the onset of rehabilitative training (week 1), and at the completion of rehabilitation (week 3). A two-way repeated measures ANOVA was performed to compare staircase pellet retrieval with the contralateral paw at the different time points and between groups. The analysis revealed that there was indeed a main effect of time on pellet retrieval in both the mass (F (1.478, 22.17) = 45.60, p <0.0001), and precision practice paradigms (F (1.107, 13.28) = 10.11, *p* = .0060), (**Figure 17A**). Following stroke, rats reached approximately 7 pellets less (or 62% of baseline) in the mass practice paradigm and 5 pellets (or 70% of baseline) in the precision practice cohort. Post-hoc test showed that change in pellets retrieved between pre-stroke and pre-rehabilitation was significant for all mass practice groups (4 sessions: *p* = .0466; 8 sessions: *p* = .0110; 12 sessions: *p* = .0105, Tukey's test). However, no significant recovery of skilled motor function was observed, as pellet retrieval on average only improved by 1.4 pellets. In the precision practice paradigm, on the other hand, the effect of time was revealed between pre- and post-rehabilitation time points in the groups with 8 (*p* = .0132, Tukey's test) and 12 sessions (*p* = .0367, Tukey's test). Thus, the observed pellet retrieval impairments immediately after stroke were not significant.

*3.8.2 Skilled reaching ability and movement practice*

The degree of forelimb impairment might affect the rat's ability to perform the rehabilitative training task. To gain a better understanding of how initial impairment influences task participation and subsequent recovery, we split the rats of both cohorts into two groups based on their staircase results pre-rehabilitation and thus eliminated the factor of training sessions. This created "low" and "high" pellet retrieval groups, separated at the cohort's medium. Correlation analyses showed that a low staircase score before training positively correlated with the total pellets retrieved (g) in rehabilitation, although not significant (**Figure 17B, left;** mass: n = 18, r = 0.58, *p* = .0991; precision: n = 15, r = 0.64, *p* = .1223**)**. Rats that were able to retrieve > 11.5 pellets in the staircase task at week 1 showed a neutral relationship with performance in rehabilitation (mass: r = -0.11, *p* = .7831; precision r = -0.01, *p* = .9777). This suggests that above a certain threshold, animals will show similar performance during rehabilitation, as measured by pellet consumption, regardless of the initial impairment. Our dataset demonstrated neutral relationships between the amounts of pellets retrieved and changes in staircase outcome before and after rehabilitation (**Figure 17B, middle**). This is in line with the fact that improvements in the staircase task were small.

*3.8.3 Infarct volume and change in reaching ability*

Correlation analyses between lesion volume ($mm^3$) and change in staircase performance (post – pre-rehabilitation) were performed as well (Figure 17B, right). No significant correlations were found, and thus we describe the size and direction of Pearson's *r.* In the mass practice paradigm, rats with a low staircase score (i.e. greater impairment) demonstrate a slight positive relationship between lesion volume and change in pellets (r = 0.33, $p$ = .3876). However, mass practice rats with a high staircase score, have a slight negative relationship (r = 0.43, $p$ = .2107). This suggests that rats with a greater impairment as measured on the staircase test and by infarct volume, will show an improvement in skilled reaching function after rehabilitation. However, rats that were less impaired initially showed smaller improvements, and skilled reaching deteriorated in some rats with a larger infarct volume, causing a negative trend. In the precision reaching task, the group with a greater impairment showed a neutral relationship (r = -0.10, $p$ = .8259). Animals with a lesser impairment demonstrated a slight positive relationship between lesion volume and change in staircase task (r = 0.26, $p$ = .0537). This suggests that after the rehabilitation paradigm, skilled reaching ability improves more for rats with a greater infarct volume, however improvement plateaus above a lesion volume of ~ 75 $mm^3$.

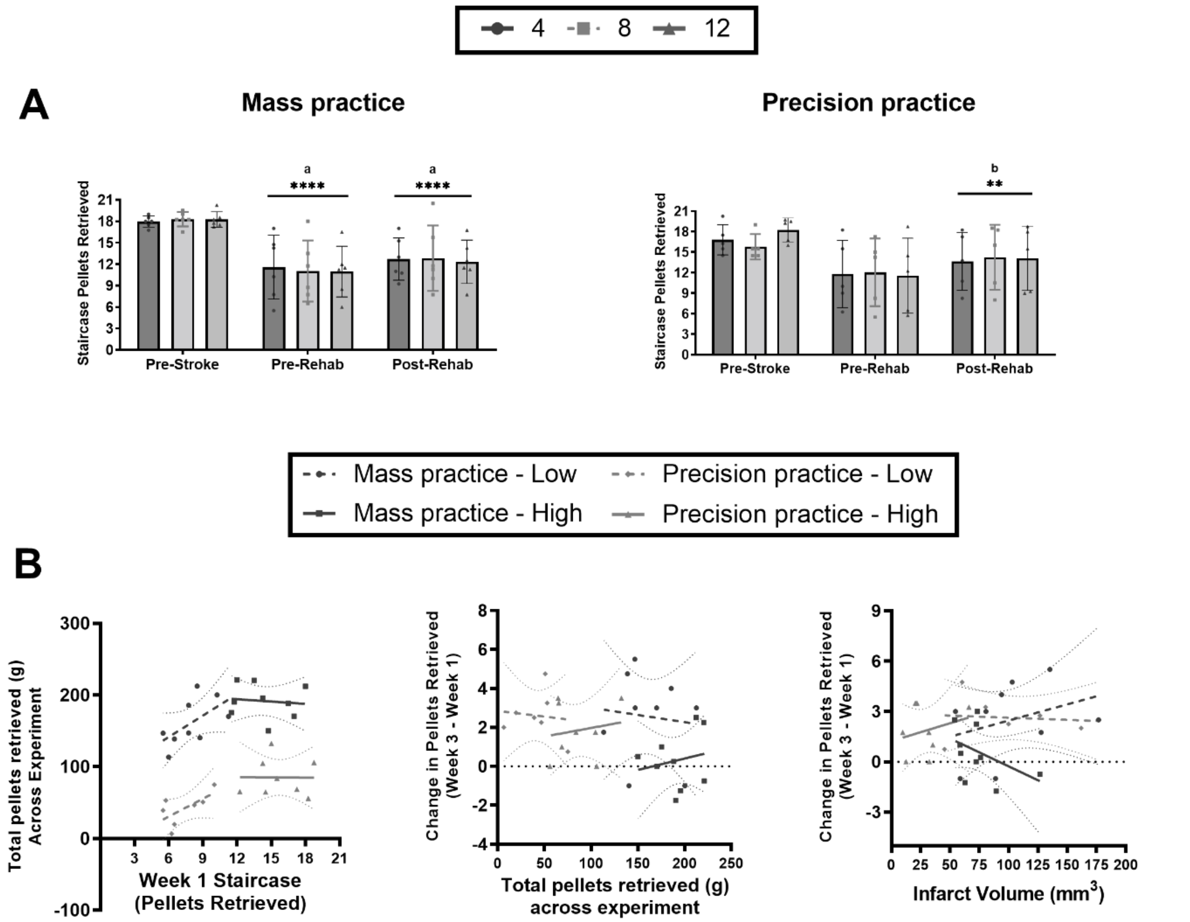


**Figure 17. Staircase performance pre-stroke, pre-rehab, and post-rehab**. Skilled reaching with the contralateral paw was assessed in the Montoya staircase task over 4-6 trials at three time points: pre-stroke, pre-rehabilitation (4 to 6 days post-stroke), and post-rehabilitation (21 to 23 days post-stroke). **A)** Two-way ANOVA showed a main effect of time in both the mass and precision practice **B)** Median split regression analyses. Animals were grouped by 'low' or 'high' number of pellets retrieved in the staircase assessment pre-rehabilitation (~week 1 post-stroke). Dashed lines indicate the 95% confidence intervals of the regressions. a, different from pre-stroke, b, different from pre-rehabilitation, ****p<0.001, ** $p < .01$.

# 4. DISCUSSION

Many stroke survivors rely on rehabilitative therapy to improve their upper limb recovery outcome after stroke. However, there are no clear guidelines as to how much physical therapy is required to achieve the desired outcome. One contributing factor is that there are no scientific standards for reporting and investigating dose in neurorehabilitation (Lang, Lohse, and Birkenmeier 2015). In preclinical studies, rehabilitative dose is typically reported as the number of sessions or successful reaches (i.e., pellet consumption). The large-scale derivation of precise dose metrics via video analysis in rodent studies has hitherto been hampered by limited human capacity for manual analysis and inadequate tools for automated analysis. This study sought to (1) automatically quantify video-derived metrics of dose (number of reaches, time on task, and paw path length) using DeepLabCut as a tool for marker-less tracking of body parts, (2) determine whether increasing dose sessions corresponds to increased forelimb use across rehabilitation and, (3) determine the relationship between video-derived dose metrics and pellet consumption. We delivered two types of forelimb training (mass or precision) to groups of rats with a varying number of daily sessions (4, 8, or 12) for 10 days. The developed algorithms could accurately detect reaches and were able to quantify time on task and calculate path length of the paw. Analyzing all training sessions with the algorithms showed that escalating the number of training sessions does not accumulate in more forelimb use. However, forelimb training at a lower frequency (4 sessions) resulted in higher intensity forelimb use within sessions, indicated by more pellet consumption, reaches, time on task, and longer paw path lengths per individual session. Video-derived metrics in precision practice correlated positively with pellets retrieved by weight. In mass practice, time on task and path length showed a neutral relationship with pellet consumption.

## 4.1 Algorithms can quantify metrics of dose

Behavioral analysis of laboratory animals is important for understanding brain function (Krakauer et al. 2017), but it has a few drawbacks. Although many behavioural assays are performed according to standardized protocols, there is still a degree of subjective interpretation by the experimenter, contributing to inter-rater variability (Mah et al. 2021). Furthermore, video analyses are often tedious to carry out, increasing the likelihood of errors with fatigue. Software packages that assist or completely replace experimenter's manual evaluation have been increasingly used over the last two decades (Pham et al. 2009; Gomez-Marin et al. 2012; Ohayon et al. 2013; Ben-Shaul 2017). The recently released DeepLabCut tool stands out from other packages as it does not require the placement of fiducial markers, has been used extensively within the research community and is inherently adaptable to a broad range of experimental needs. In this study, we investigated if DLC's automated pose estimation can be used to quantify specific metrics of dose that have been proposed to be important for identifying therapeutic rehabilitation (Hayward et al. 2021).

We developed algorithms to accurately estimate the number of reaching movements in two forms of rehabilitation: a precision reaching task and a trough reaching task. Validation was done by comparing reaching events identified automatically with frame-by-frame manual inspection. The first version of the algorithm, designed using Tracker output, detected 90.2% of the reaching movements in the mass practice task (**Table 1**). After refinement and modification of our DLC analysis pipeline, the final version of the algorithm detected up to 98.8% of the reaches in the mass practice task and 81.7% percent of the precision task reaches. Despite this high agreement, the mass practice algorithms detected a statistically significant different number of reaches when calculating the median of a sample (**Figure 6**). Although this was unexpected, it indicates there is a consistent overestimation in the number of reaching events by our algorithm relative to manual analysis. This is due to false positive events that are identified in addition to the near-perfect detection of true reaching events. Movements underlying false positive events are for example slight lowering of the paw after placing the pellet into the mouth or placing the paw along the side of the platform to move forward. In future experiments mistakes like these could be circumvented by more stringent boundary conditions in the formula that describes reaches.

The formulas for reaching movements are based on basic kinematics (i.e. paw trajectory). These, and other kinematic variables have been previously used to characterize reaching movements in healthy and injured rodents (Braun, Andrews, and Kartje 2012; Lai et al. 2015; Balbinot et al. 2018; Becker et al. 2020), and some studies applied 3D kinematics (Azim et al. 2014; Guo et al. 2015; Bova et al. 2020). One study used an image processing program to assess end point and kinematic features of skilled reaching in rats before and after M1 lesion (Nica et al. 2018). Top view recordings of the reaching area were used to capture the paw interaction with the pellet. Their algorithm detected the reach, grasp, and retracting features of trials, whereby the end of a reach was determined by the maximum extension of the forelimb. These features were then used to determine the outcome of a trial (success, miss, slip, or drop). Overall, the labeling algorithm had a sensitivity of 86% pre-lesion and 92% following motor cortex injury.

Another group recently implemented 3D kinematics using DLC, to analyze up to 4 weeks of recovery from focal cerebral ischemia (Weber et al. 2021), however in contrast to our reaching analysis, this study performed gait analysis in mice. The authors demonstrated compensatory strategies by alterations in the ipsilateral trajectories. Furthermore, tests that were assisted with DLC analysis were benchmarked to conventional behavioral set-ups. This revealed significantly reduced analysis time while minimizing observer biases during the evaluation.

Together, these data show that computer vision techniques can accurately assess and qualify rodent behavior before and after motor injury. This is an important point, as one challenge inherent to both automated and manual kinematic analysis is that following injury the behavior of interest may be modified to a degree that makes it difficult to apply the same scoring criteria as before the injury. For example, within reaching tasks, entire segments of the reaching movement may be absent or severely modified (Moon et al. 2009), thus potentially altering the efficacy of automated identification of behavioral events. Our algorithm was built based on post-stroke behavioral data and uses a minimal displacement rule to detect reaches. We did not assess the performance of the algorithm on pre-stroke data, however, we observed a greater number of reaching events pre-stroke along with higher pellet retrieval rates. It could be possible that the increase is caused by false positive observations if there are systemic deviant movement patterns that fit the criteria of the algorithm. However, we assume this to be unlikely for our model that seeks primarily forward movements. Furthermore, reaches that do not meet the algorithms' criteria (i.e. are below the threshold) would be categorized as false negative and therefore not contribute to the observed increase. Thus, although there may be discrepancies in the performance of the algorithm on pre vs. post-stroke reaching behavior, we assume that these alterations are not significant.

### 4.2 Behavioral analysis time can effectively be reduced through subsampling

Video is a powerful medium to capture animal behavior. It allows observation of subtle movements undetectable by the human naked eye, scoring of behavior at a slower playback speed, and reconstructing movement trajectories in 3D with the use of multiple cameras. However, frame-by-frame video analysis is extremely time-consuming and requires significant experimenter training to perform well. In our mass practice data set, it cost 45 hours to track two body parts in a 15-minute video. Others have also commented on the time-consuming analysis of behavioral assays (Anderson and Perona 2014; Weber et al. 2021). We assessed how subsampling techniques may reduce analysis time whilst maintaining an accurate estimation of the measured variable. This was done by computing the percent error relative to the authentic value after sampling with various subsample lengths and frequencies. Our results showed that when short bits of footage are analyzed frequently, the outcome measure can be estimated with an accuracy of up to +/- 5% error (**Figure 7C-D)**. Specifically, in a selection of our data set, we showed that determining group averages by analyzing 16% of the videos, (i.e., 10 s/1 min, equally spaced across the video length), gives a similar result as analyzing 100% (**Figure 8**). This is not surprising as subsampling methods are commonly applied in different fields of science. Furthermore, we found that rat's reaching activity tapers off toward the end of a training session. Understanding how behaviors are distributed over time is important to make an informed decision on how to collect representative subsamples. Our data demonstrated that when one wants to analyze a subsample of the video footage to predict the overall behavior, this sample should originate from multiple time points across the video to capture both occasions of higher and lower activity.

A similar subsampling method to ours has been used to characterize physical activity patterns of stroke patients in an acute stroke unit (Bernhardt et al. 2004). In this study, observers recorded stroke patients' activities in a medical ward every 10 minutes throughout the day. With observations lasting ~1 minute every 10 minutes, approximately 10% of the patient's physical activity was logged. Although this study did not benchmark their approach, with the results from our subsampling analysis, we know that this is a reliable method of arriving at an estimation of overall behavior, provided that all behaviors are captured in the subsample.

In conclusion, subsampling data is a well-known method to reduce analysis time and retain accuracy. Our results show that this method is also effective in rat reaching tasks, and presumably, it can be applied in other tasks such as locomotion. Experimenters may want to consider this approach when working with large datasets and needing to expedite video analysis.

### 4.3 Escalating session dose does not result in more forelimb use

There is a general understanding that large quantities of motor practice are a key component to induce cortical reorganization and improved behavioral function (Kleim and Jones 2008), (Nudo and Milliken 1996). Furthermore, a positive dose-response relationship was found in a meta-analysis of randomized controlled trials (Lohse, Lang, and Boyd 2014). Preclinical studies in rodents also demonstrated improved outcomes in motor function if a higher dose (i.e. sessions, pellets retrieved, or number of reaches) was received (Bell et al. 2015; MacLellan et al. 2011; Jeffers et al. 2018). However, it remains unclear how maximal forelimb use can be evoked through rehabilitative training. Does forelimb use increase proportionally with the number of training sessions? To answer this question, we systematically varied the number of daily training sessions in the mass and precision practice paradigm. We measured multiple metrics of dose, both at the level of session intensity (pellets retrieved, number of reaches, paw path length) and session density (time on task), as described in the nonpharmacological dose articulation framework (Hayward et al. 2021). No statistically significant differences in the grand sum of pellets retrieved (g), number of reaches, time on task (s), and paw path length (cm) were observed after the rehabilitation paradigms. Although, in the precision practice paradigm, rats with 8 daily sessions tended to have higher accumulated doses than rats with 4 or 12 sessions (**Figure 14**). These results do not support the hypothesis that dose increases proportionally with the number of rehabilitative training sessions provided.

Interestingly, the reason why we did not observe a proportional increase in the summed values is that animals became increasingly less active as the number of daily training sessions increased. (See **Figure S 3** for additional data on pellet retrieval throughout rehabilitation days). This difference between groups was significant in the mass practice paradigm **(Figure 13**). In the precision practice paradigm, we observed a similar pattern in most outcome measures, (except time on task), although not significant (**Figure 15)**. In total, rats with 4 daily sessions were able to grasp ~4 g, or 88 pellets in 15 minutes, accumulating to ~352 pellets per day. Previous preclinical studies operating *ad libitum* reaching paradigms demonstrated similar results. Rats that were given 20 minutes access to 4.5 g of pellets on a tray, grasped 3.5 g on the first days of rehabilitation after ischemic sensorimotor cortex stroke, and up to 4.5 g after 5 weeks of training (Maldonado et al. 2008). Voluntary or restricted access seems a key factor in the number of pellets rats will retrieve. In a study by MacLellan et al. (2011), rats had access to a reaching apparatus for 4 hours per day and consumed 250 pellets. Voluntary access provided over 3x 15-minute bouts resulted in 300 reaches, whereas animals that could not leave the apparatus performed up to 600 reaches per day (Jeffers et al. 2018).

Together, the results of these studies suggest that rats will retrieve more pellets when reintroduced to the device multiple times and access is not voluntary. In addition, our results show that a shorter rest period has an unfavorable impact on training performance. Rats with 8 and 12 sessions rested for 30 and 15 minutes between sessions, respectively, and tended to be less active during training than rats with 75 minutes rest and 4 sessions (also see **Figure S 5**). The underlying cause of this is not clear. Satiation is probably not a limiting factor on its own, as the phenomenon also occurs in precision reaching task, where fewer pellets are consumed. Another factor at play could be lower water consumption between sessions as diminished water intake can reduce food intake (Hsiao and Lloyd 1969). Furthermore, repeated exposure to behavioral tests has been shown to alter behavior. Brown and Nemes (2008) observed differed frequency of head-dipping in the hole-board task in rats, across 10 trials. However, this was attributed to a decrease in fearfulness. In contrast, repeated daily exposure in the elevated zero-maze, enhanced anxiety in mice (Cook, Crounse, and Flaherty 2002). Compared to the elevated zero-maze, however, we do not suspect our training apparatus to be anxiety-inducing and the animals should habituate more easily. Nonetheless, frequent transfer between training and homecage environment may be stressful to the animal.

### 4.4 Relationship between video metrics and pellets retrieved

Until now, preclinical studies have mostly reported dose as pellets retrieved (either by weight or by number) or number of sessions (Biernaskie and Corbett 2001; Biernaskie, Chernenko, and Corbett 2004; Maldonado et al. 2008; MacLellan et al. 2011; Bell et al. 2015; Jeffers et al. 2018). For that reason, pellet retrieved can be considered a gold standard for the measurement of dose at the level of intensity. One of the ideas behind regarding rehabilitation dose as multidimensional is that it provides a better picture of what was delivered during a non-pharmacological intervention (Hayward et al. 2021). Our dataset allows us to compare these additional variables with the gold standard. What is the association between our dose metrics derived from video analysis and the weight of the pellets retrieved? We observed strong positive correlations between video-derived metrics of dose and pellets retrieved in the precision practice task (**Figure 16**). In contrast, in the mass practice paradigm, only the number of reaches had a significant positive correlation with pellets retrieved (**Figure 16**). Time on task and paw path length showed a weak positive correlation, which was not significant.

The associations between the number of reaches and pellets retrieved were predicted, as they are both a measure of rehabilitation intensity. In that same vein, path length was also expected to associate with pellets retrieved. However, time on task was not expected to relate to pellets retrieved. Previous work has demonstrated that session intensity can be quite variable, even if session time remains equal. Rats that were given access to a mass practice reaching apparatus 3x daily for 15 minutes following stroke demonstrated substantial variability in the number of performed reaches, ranging from ~300 to 600 (Jeffers et al. 2018). However, the conditions under which access was provided were unequal: rats with voluntary access performed fewer reaches than rats that could not leave the apparatus. Despite equivalent session times, it is unknown how the conditions affect the amount of time the animals were actively engaged. It underlines the notion that solely reporting session time, is insufficient in capturing the delivered dose. Similarly, highly variable number of exercise repetitions were counted in stroke patients within a fixed, 30 minute, session length, ranging from 4 up to 369 (Scrivener et al. 2011). These studies do not tell us whether fewer repetitions were completed because less time was spent on task, or in spite a similar time on task.

The contrasts in mass and precision practice associations between time on task and pellets retrieved is attributable to a discrepancy in how time on task was measured. In the mass practice, time on task was exclusively based on the location of the nose, relative to the trough. Specifically, a rat was on task if the nose was oriented towards the pellets. However, it was possible for rats to elevate their head, no longer being on task, while they perform a reach. In precision practice, the equation for time on task incorporated multiple parameters, including the paw's position. This was necessary because the nose was not continuously visible, as rats may interact with the pedestal while rearing on their hindlimbs. Thus, the discrepancy lies in that in mass practice a reach could occur while the rat is not on task whereas in precision practice the rat is always on task while reaching.

Interestingly, we did not find a positive correlation between path length and pellets retrieved in mass practice. This, in combination with the observation that the number of reaches was positively associated with pellets retrieved, suggests that there is a disparity in reaching strategies among animals. It is possible that rats that retrieved more pellets, were less injured and demonstrated a more efficient strategy, consisting of shorter and smoother movements. In contrast, rats with greater impairments may have adopted a strategy resulting in longer, jerkier trajectories. Kinematic analyses in rodents post-stroke have revealed elongation and reduced smoothness of the reach trajectory in the single pellet and staircase reaching task (Braun, Andrews, and Kartje 2012; Lai et al. 2015; Balbinot et al. 2018). In our precision reaching paradigm reaching trials were often contaminated with a hovering type of movements and occasionally reaching attempts with the other paw. This form of less indisputable reaching movement could mask the detection of subtle discrepancies in reaching strategies. Instead, the increased accumulative path length is potentially a more accurate indication for more reaching attempts.

### 4.5 Rehabilitation was not effective in improving outcome

The goal of rehabilitative therapy following stroke is to restore lost functions and reduce acquired impairments. Standardized behavioral tests are designed to assess specific sensorimotor functions. Forelimb function is commonly gauged through reaching tasks (Schaar, Brenneman, and Savitz 2010). In this study skilled contralesional reaching ability was assessed at three key time points, before stroke, pre-and post-rehabilitation, using the Montoya staircase task. The rats in the mass practice experiment demonstrated significant impairments after photothrombotic stroke. However, lost function was not recovered after rehabilitation. Rats in the precision practice paradigm showed a reduction in staircase pellets retrieval, however not statistically significant compared to pre-stroke. In contrast to the mass practice, rats with 8 and 12 sessions per day demonstrated significantly improved reaching ability after rehabilitation. Thus, in most animals, two weeks of reaching practice did not enhance recovery. This is not very surprising considering that the study was not designed to improve recovery per se. Rather, the main aim was to investigate dose escalation and capture the multidimensionality of rehabilitative training. However, significant improvements in motor function have been observed after 2-4 weeks of behavioral training.

In Nudo and colleagues' classic experiment demonstrating cortical reorganization of the upper limb representation in non-human primates after small lesions in the M1 hand area, the animals were trained in a skilled reaching task, for 3-4 weeks (Nudo and Milliken 1996). After 2 weeks of daily tray reaching training (20 minutes), rats with ET-1 induced SMC damage achieved a higher success rate in the single pellet reaching task compared to untrained animals (Maldonado et al. 2008). After 4 weeks of training, the success rate maintained above baseline levels. A similar result on the staircase task (with the contralesional limb) was found in rats exposed to EE in combination with reach training (Biernaskie and Corbett 2001). The authors suggested that neural plasticity in the form of enhanced dendritic complexity and length in the non-injured hemisphere contributed to the success of the rehabilitation paradigm. Daily access to a reaching apparatus during the dark cycle after MCAO damage also improved contralateral forelimb reaching success on the staircase task (MacLellan et al. 2011). Two weeks of training resulted in a significant improvement compared to groups with limited or no training, and this difference was sustained until the end of the 5-week rehabilitation period. Together, these outcomes suggest that two weeks of behavioral training may be sufficient to detect recovery of motor function. Furthermore, it has been suggested that a higher daily training frequency expedites recovery (Bell et al. 2015). In accordance with this concept, our animals would have the potential to show recovery following this paradigm. However, there is one main reason why recovery was not detected in all rehabilitation groups of this study.

The photothrombotic protocol induced strokes that resulted in mostly mild impairments. Only 20% of all rats demonstrated contralesional forelimb reaching impairment at week 1 post-stroke equal to what was shown in previous rehabilitation studies, i.e., 30% of baseline score (Biernaskie and Corbett 2001; MacLellan et al. 2011; Jeffers et al. 2018). Taking the proportional recovery rule into account, (Jeffers, Karthikeyan, and Corbett 2018; Jeffers et al. 2018), the majority of rats in this study are subjected to a ceiling effect. There is only a small range of improvement possible if the right training intensity is met. This phenomenon is reflected in the mass practice data points in **Figure 17B**, middle panel. Nearly all rats show total pellets retrieved in a range from 150 – 225 g. However, despite a similar amount of pellets retrieved the less impaired animals demonstrate approximately no improvement. This could be due to the fact that the training was not intense enough in these animals to promote recovery. It contrasts the less impaired animals in precision reaching, do show some improvement. They performed more reaches during rehabilitation and therefore experienced more intense training. However, this does not fully explain why more severely impaired rats showed similar improvements in the staircase task, despite that their training intensity was likely lower. One possible explanation is that their skilled hand use improved through spontaneous recovery. The ambiguity of these results warrants further research into the effect of these upper limb training paradigms on recovery and to elucidate the discrepancies between the regimens. We believe the paradigms have the potential to induce significant recovery, in a cohort with less variable and more severe impairments.

**4.6 Limitations & future directions**

This study has some weaknesses that should be considered when interpreting the results. A unique aspect of this project was the ability to, in both experiments, deliver rehabilitative training to the entire cohort simultaneously. Unfortunately, as a consequence, the physical capacity of our experimental setup limited the group sizes. As such, statistical power remains low. Moreover, unexpected intra-group variability in the precision practice paradigm prevents drawing strong statistical conclusions from that experiment. The most pronounced reason for uneven performance is the wide variability of lesion volume and location, particularly in the precision practice experiment. Lesions were either very small $< 50\ mm^3$, or very large, $> 125\ mm^3$. Furthermore, the bigger infarcts extended from -3.0 to +6.3 anterior-posterior relative to bregma. Consequently, impacting hindlimb function and frontal association cortex as well. The female sex of rats in this study may also have impacted infarct volume. The effects of the estrous cycle in female rats on lesion volume have been documented (Alkayed et al. 1998; Liao et al. 2001). In one rehabilitation study, experimenters investigated estrous stage effects at the time of ET-1 lesion induction and found that when estrogen and progesterone levels were at their peak, rats had more cortical volume remaining in the SMC region (Maldonado et al. 2008).

Another aspect to consider is we conducted our study in young and healthy animals. Although the effects of rehabilitation on recovery may be different in aged animals (with comorbidities) (Alaverdashvili and Whishaw 2010; Brown, Marlowe, and Bjelke 2003; Buga, Di Napoli, and Popa-Wagner 2013). It is unlikely that the ratio between the number of training sessions in the different outcome measures will change substantially. Nonetheless, translational value would increase if this were investigated. A final remark regarding the animals in this study concerns the absence of histological analysis for structural changes or biomarkers of plasticity. Future preclinical experiments aiming to elucidate the role of different dose metrics in upper limb recovery should incorporate measures of neuroplasticity. Thereby, it is recommended to investigate this in animals with moderate-severe impairments to be able to observe training-related improvements.

There are two main limitations with regards to experimental design and behavioral analysis. Not only was the number of training sessions manipulated, so was the resting time between sessions. As mentioned earlier, this was necessary to standardize the amount of time the animals are in the experimental environment. Our data show, however, that this impacts performance in individual training sessions, resulting in equal amounts of total delivered dose. As such, one must keep in mind that our findings are limited to the effect of training frequency on movement repetitions. Moreover, because the varied number of training sessions elicits similar amounts of forelimb use, we cannot address which number of training sessions best promotes recovery. To exclusively assess the effect of number of sessions on movement repetitions and recovery, experimenters would need to maintain resting time as a constant parameter. Further, precision reaching behavior was not shaped through a stringent protocol as usual in the single pellet reaching task (Whishaw and Pellis 1990). This left the animals to expose a more liberate reaching style. The disadvantage of this, in combination with the cyclic presentations of pellets, is that rats can exhibit a hovering arm movement pattern in anticipation of the target's appearance. These jerky, aimless movements, are difficult to distinguish from normal reaching movements in both visual frame-by-frame analysis and mathematical algorithms. Furthermore, it was possible for the rats to extend their ipsilesional arm through the front slit, yet without being able to grab the pellet. Unfortunately, it can be challenging to discriminate which paw is passed through the reaching slot on some of the video recordings. Consequently, it is possible for the neural network to occasionally track the wrong paw. Nonetheless, we demonstrated that our algorithm has a high degree of sensitivity and specificity relative to manual scoring. On a final note, our algorithms are designed to capture movements in the sagittal plane (2D) where our reference points were positioned. Hence, movements (distances) outside this plane, i.e. towards or away from the camera, are not accurately perceived. However, because forelimb movements occurred mostly in the same plane, these distortions are negligible. Future experiments may circumvent some of the behavioral analysis issues by restraining ipsilesional forelimb reaching attempts with a bracelet (Whishaw, O'connor, and Dunnett 1986) and using higher framerate video recordings to better capture the fast movements and improved characterization of hovering versus reaching movements.

# 5 CONCLUSION & SIGNIFICANCE

Until recently, there were no clear guidelines for how to report rehabilitation dose in stroke recovery research. This lack of standardization is a limiting factor in understanding optimal rehabilitation dose and its active ingredients. Preclinical research plays an important role in this discovery pipeline. In our study, we aimed to utilize a state-of-the-art video analysis tool to measure metrics of dose that are in alignment with the latest guidelines. Additionally, we aimed to investigate the effects of dose escalation on training performance in two upper limb rehabilitative training tasks and elucidate the relationship between the novel and traditional metrics of dose.

We showed that with the use of DeepLabCut software, rat reaching behavior can accurately be classified by computer algorithms. This result contributes to a growing body of work implementing artificial intelligence systems to automate behavioral analysis. These systems remove the laborious burden of manual analysis, promoting the conduction of large behavioral experiments with subsequent video analysis of multiple variables.

Our finding that increasing dose sessions does not produce more forelimb movement or time on task requires careful interpretation. We found an inverse relationship between number of training sessions and work within individual sessions. This shows that rehabilitative training at a frequency that is too high may negatively impact performance per session. As a result, a low, medium, or high frequency of sessions will give a similar upper limb training dose across groups. However, for clinical translation, it is important to know that rats have different motivational triggers than humans. Nonetheless, patients may also experience several factors that affect their training performance, such as general fatigue, or discouragement by impairments.

We are the first to quantify multiple components of rehabilitation dose within a preclinical study. We demonstrated that the relationships between video-derived metrics and the conventional metric of dose (retrieved pellets) are different for precision and mass practice reaching tasks. The exact interpretation remains unclear, but it indicates discrepancies between the two tasks. Future work should explore the contributions of these metrics on recovery in animal models with moderate-severe impairments. Furthermore, examination of neural tissue is required to elucidate the neural mechanism underlying "optimal" dose. Together, this will provide mechanistic evidence that can aid in early phase clinical trial design.

## REFERENCES


Adkins, Deanna L, Jeffery Boychuk, Michael S Remple, and Jeffrey A Kleim. 2006. "Motor Training Induces Experience-Specific Patterns of Plasticity across Motor Cortex and Spinal Cord." *J Appl Physiol* 101: 1776–82. https://doi.org/10.1152/japplphysiol.00515.2006.-The.

Alaverdashvili, M., S. K. Moon, C. D. Beckman, A. Virag, and I. Q. Whishaw. 2008. "Acute but Not Chronic Differences in Skilled Reaching for Food Following Motor Cortex Devascularization vs. Photothrombotic Stroke in the Rat." *Neuroscience* 157 (2): 297–308. https://doi.org/10.1016/J.NEUROSCIENCE.2008.09.015.

Alaverdashvili, M, and I Q Whishaw. 2010. "Compensation Aids Skilled Reaching in Aging and in Recovery from Forelimb Motor Cortex Stroke in the Rat." *NSC* 167: 21–30. https://doi.org/10.1016/j.neuroscience.2010.02.001.

Alkayed, Nabil J., Izumi Harukuni, Alane S. Kimes, Edythe D. London, Richard J. Traystman, and Patricia D. Hurn. 1998. "Gender-Linked Brain Injury in Experimental Stroke." *Stroke* 29: 159–66. https://doi.org/10.1161/01.STR.29.1.159.

Amki, M. El, P. Baumgartner, O. Bracko, A. R. Luft, and S Wegener. 2017. "Task-Specific Motor Rehabilitation Therapy After Stroke Improves Performance in a Different Motor Task: Translational Evidence." *Translational Stroke Research* 8 (4): 347–50. https://doi.org/10.1007/S12975-016-0519-X.

Anderson, David J., and Pietro Perona. 2014. "Toward a Science of Computational Ethology." *Neuron* 84 (1): 18–31. https://doi.org/10.1016/J.NEURON.2014.09.005.

Azim, Eiman, Juan Jiang, Bror Alstermark, and Thomas M. Jessell. 2014. "Skilled Reaching Relies on a V2a Propriospinal Internal Copy Circuit." *Nature 2014 508:7496* 508 (7496): 357–63. https://doi.org/10.1038/nature13021.

Bacigaluppi, Marco, Giancarlo Comi, and Dirk M. Hermann. 2010. "Animal Models of Ischemic Stroke. Part Two: Modeling Cerebral Ischemia." *The Open Neurology Journal* 4 (2): 34. https://doi.org/10.2174/1874205X01004020034.

Balbinot, Gustavo, Clarissa Pedrini Schuch, Matthew S Jeffers, Matthew W Mcdonald, Jessica M Livingston-Thomas, and Dale Corbett. 2018. "Post-Stroke Kinematic Analysis in Rats Reveals Similar Reaching Abnormalities as Humans." *Scientific Reports* 8: 1–13. https://doi.org/10.1038/s41598-018-27101-0.

Barbay, Scott, Erik J. Plautz, Kathleen M. Friel, Shawn B. Frost, Numa Dancause, Ann M. Stowe, and Randolph J. Nudo. 2006. "Behavioral and Neurophysiological Effects of Delayed Training Following a Small Ischemic Infarct in Primary Motor Cortex of Squirrel Monkeys." *Experimental Brain Research* 169 (1): 106–16. https://doi.org/10.1007/s00221-005-0129-4.

Becker, Matthew I., Dylan Calame, Julia Wrobel, and Abigail L. Person. 2020. "Quantification of Mouse Reach Kinematics as a Foundation for Mechanistic Interrogation of Motor Control." *BioRxiv*, April. https://doi.org/10.1101/2020.04.24.060533.

Bell, Jared A., Malerie L. Wolke, Ryan C. Ortez, Theresa A. Jones, and Abigail L. Kerr. 2015.

“Training Intensity Affects Motor Rehabilitation Efficacy Following Unilateral Ischemic Insult of the Sensorimotor Cortex in C57BL/6 Mice.” *Neurorehabilitation and Neural Repair* 29 (6): 590–98. https://doi.org/10.1177/1545968314553031.

Ben-Shaul, Yoram. 2017. “OptiMouse: A Comprehensive Open Source Program for Reliable Detection and Analysis of Mouse Body and Nose Positions.” *BMC Biology* 15 (41). https://doi.org/10.1186/S12915-017-0377-3.

Bernhardt, Julie, Karen Borschmann, Lara Boyd, S Thomas Carmichael, Dale Corbett, Steven C Cramer, Tammy Hoffmann, et al. 2016. “Moving Rehabilitation Research Forward: Developing Consensus Statements for Rehabilitation and Recovery Research:” *International Journal of Stroke* 11 (4): 454–58. https://doi.org/10.1177/1747493016643851.

Bernhardt, Julie, Helen Dewey, Amanda Thrift, and Geoffrey Donnan. 2004. “Inactive and Alone: Physical Activity within the First 14 Days of Acute Stroke Unit Care.” *Stroke* 35 (4): 1005–9. https://doi.org/10.1161/01.STR.0000120727.40792.40.

Biernaskie, Jeff, Garry Chernenko, and Dale Corbett. 2004. “Efficacy of Rehabilitative Experience Declines with Time after Focal Ischemic Brain Injury.” *Journal of Neuroscience* 24 (5): 1245–54. https://doi.org/10.1523/JNEUROSCI.3834-03.2004.

Biernaskie, Jeff, and D. Corbett. 2001. “Enriched Rehabilitative Training Promotes Improved Forelimb Motor Function and Enhanced Dendritic Growth after Focal Ischemic Injury.” *Journal of Neuroscience* 21 (14): 5272–80. https://doi.org/10.1523/jneurosci.21-14-05272.2001.

Birkenmeier, Rebecca L., Eliza M. Prager, and Catherine E. Lang. 2010. “Translating Animal Doses of Task-Specific Training to People with Chronic Stroke in 1-Hour Therapy Sessions: A Proof-of-Concept Study.” *Neurorehabilitation and Neural Repair* 24 (7): 620–35. https://doi.org/10.1177/1545968310361957.

Bova, Alexandra, Krista Kernodle, Kaitlyn Mulligan, and Daniel Leventhal. 2020. “Automated Rat Single-Pellet Reaching with 3-Dimensional Reconstruction of Paw and Digit Trajectories.” *J Vis Exp.*

Braun, Robynne G., Ellen M. Andrews, and Gwendolyn L. Kartje. 2012. “Kinematic Analysis of Motor Recovery with Human Adult Bone Marow Derived Somatic Cell Therapy in a Rat Model of Stroke.” *Neurorehabilitation and Neural Repair* 26 (7): 898–906.

Broeks, J. G., G. J. Lankhorst, K. Rumping, and A. J.H. Prevo. 1999. “The Long-Term Outcome of Arm Function after Stroke: Results of a Follow-up Study.” *Disability and Rehabilitation* 21 (8): 357–64. https://doi.org/10.1080/096382899297459.

Brown, Allen W., Kimberly J. Marlowe, and Börje Bjelke. 2003. “Age Effect on Motor Recovery in a Post-Acute Animal Stroke Model.” *Neurobiology of Aging* 24 (4): 607–14. https://doi.org/10.1016/S0197-4580(02)00129-X.

Brown, Craig E., Ping Li, Jamie D. Boyd, Kerry R. Delaney, and Timothy H. Murphy. 2007. “Extensive Turnover of Dendritic Spines and Vascular Remodeling in Cortical Tissues Recovering from Stroke.” *Journal of Neuroscience* 27 (15): 4101–9. https://doi.org/10.1523/JNEUROSCI.4295-06.2007.

Brown, Gillian R., and Christopher Nemes. 2008. "The Exploratory Behaviour of Rats in the Hole-Board Apparatus: Is Head-Dipping a Valid Measure of Neophilia?" *Behavioural Processes* 78 (July): 442–48. https://doi.org/10.1016/J.BEPROC.2008.02.019.

Buga, A. M., Mario Di Napoli, and A. Popa-Wagner. 2013. "Preclinical Models of Stroke in Aged Animals with or without Comorbidities: Role of Neuroinflammation." *Biogerontology* 14 (December): 651–62. https://doi.org/10.1007/S10522-013-9465-0/TABLES/1.

Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (3): 1–67. https://doi.org/10.18637/JSS.V045.I03.

Carmichael, S. Thomas, Ivonne Archibeque, Linslee Luke, Tim Nolan, Janneth Momiy, and Songlin Li. 2005. "Growth-Associated Gene Expression after Stroke: Evidence for a Growth-Promoting Region in Peri-Infarct Cortex." *Experimental Neurology* 193 (2): 291–311. https://doi.org/10.1016/j.expneurol.2005.01.004.

Carmichael, S. Thomas, and Marie Françoise Chesselet. 2002. "Synchronous Neuronal Activity Is a Signal for Axonal Sprouting after Cortical Lesions in the Adult." *Journal of Neuroscience* 22 (14): 6062–70. https://doi.org/10.1523/JNEUROSCI.22-14-06062.2002.

Clarkson, Andrew N., Héctor E. López-Valdés, Justine J. Overman, Andrew C. Charles, K. C. Brennan, and S Thomas. Carmichael. 2013. "Multimodal Examination of Structural and Functional Remapping in the Mouse Photothrombotic Stroke Model." *Journal of Cerebral Blood Flow and Metabolism* 33 (February): 716–23. https://doi.org/10.1038/jcbfm.2013.7.

Cook, Melloni N., Martha Crounse, and Lorraine Flaherty. 2002. "Anxiety in the Elevated Zero-Maze Is Augmented in Mice after Repeated Daily Exposure." *Behavior Genetics* 32 (2): 113–18. https://doi.org/10.1023/A:1015249706579.

Corbett, Dale, S. Thomas Carmichael, Timothy H. Murphy, Theresa A. Jones, Martin E. Schwab, Jukka Jolkkonen, Andrew N. Clarkson, et al. 2017. "Enhancing the Alignment of the Preclinical and Clinical Stroke Recovery Research Pipeline: Consensus-Based Core Recommendations From the Stroke Recovery and Rehabilitation Roundtable Translational Working Group." *International Journal of Stroke* 12 (5): 462–71. https://doi.org/10.1177/1545968317724285.

Eshkol, N., and A Wachman. 1958. *Movement Notation*. Edited by Nicholson and Weidenfeld. Londen.

Fluri, Felix, Michael K. Schuhmann, and Christoph Kleinschnitz. 2015. "Animal Models of Ischemic Stroke and Their Application in Clinical Research." *Drug Design, Development and Therapy* 9: 3445–54. https://doi.org/10.2147/DDDT.S56071.

Garcia, JH, KF Liu, and KL Ho. 1995. "Neuronal Necrosis after Middle Cerebral Artery Occlusion in Wistar Rats Progresses at Different Time Intervals in the Caudoputamen and the Cortex." *Stroke* 26 (4): 636–42. https://doi.org/10.1161/01.STR.26.4.636.

Gomez-Marin, Alex, Nicolas Partoune, Greg J. Stephens, and Matthieu Louis. 2012. "Automated Tracking of Animal Posture and Movement during Exploration and Sensory Orientation Behaviors." *PLoS ONE* 7 (8): e41642. https://doi.org/10.1371/journal.pone.0041642.

Guo, Jian Zhong, Austin R. Graves, Wendy W. Guo, Jihong Zheng, Allen Lee, Juan Rodríguez-Gonzá Lez, Nuo Li, et al. 2015. "Cortex Commands the Performance of Skilled Movement." *ELife* 4 (December): e10774. https://doi.org/10.7554/ELIFE.10774.

Haenlein, Michael, and Andreas Kaplan. 2019. "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence." *California Management Review* 61 (4): 5–14. https://doi.org/10.1177/0008125619864925.

Hayward, Kathryn S., Leonid Churilov, Emily J. Dalton, Amy Brodtmann, Bruce C.V. Campbell, David Copland, Numa Dancause, et al. 2021. "Advancing Stroke Recovery Through Improved Articulation of Nonpharmacological Intervention Dose." *Stroke* 52 (2): 761–69. https://doi.org/10.1161/STROKEAHA.120.032496.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 770–78. https://doi.org/10.1109/CVPR.2016.90.

Heart and Stroke Foundation of Canada. 2019. "5.1 Management of the Upper Extremity Following Stroke | Canadian Stroke Best Practices." 2019. https://www.strokebestpractices.ca/recommendations/stroke-rehabilitation/management-of-the-upper-extremity-following-stroke.

Hsiao, Sigmund, and Margaret A. Lloyd. 1969. "Do Rats Drink Water in Excess of Apparent Need When They Are given Food?" *Psychonomic Science* 15 (3): 155–56. https://doi.org/10.3758/BF03336254.

Humm, J. Leigh, Dorothy A. Kozlowski, Debra C. James, Jeffrey E. Gotts, and Timothy Schallert. 1998. "Use-Dependent Exacerbation of Brain Damage Occurs during an Early Post- Lesion Vulnerable Period." *Brain Research* 783 (2): 286–92. https://doi.org/10.1016/S0006-8993(97)01356-5.

Ishida, Akimasa, Sachiyo Misumi, Yoshitomo Ueda, Yuko Shimizu, Jung Cha-Gyun, Keigo Tamakoshi, Kazuto Ishida, and Hideki Hida. 2015. "Early Constraint-Induced Movement Therapy Promotes Functional Recovery and Neuronal Plasticity in a Subcortical Hemorrhage Model Rat." *Behavioural Brain Research* 284: 158–66. https://doi.org/10.1016/j.bbr.2015.02.022.

Jadhav, Ashutosh P., Shashvat M. Desai, Cynthia L. Kenmuir, Marcelo Rocha, Matthew T. Starr, Bradley J. Molyneaux, Bradley A. Gross, Brian T. Jankowitz, and Tudor G. Jovin. 2018. "Eligibility for Endovascular Trial Enrollment in the 6- to 24-Hour Time Window." *Stroke* 49 (4): 1015–17. https://doi.org/10.1161/STROKEAHA.117.020273.

Jeffers, Matthew Strider, and Dale Corbett. 2018. "Synergistic Effects of Enriched Environment and Task-Specific Reach Training on Poststroke Recovery of Motor Function." *Stroke* 49 (6): 1496–1503. https://doi.org/10.1161/STROKEAHA.118.020814.

Jeffers, Matthew Strider, Sudhir Karthikeyan, and Dale Corbett. 2018. "Does Stroke Rehabilitation Really Matter? Part A: Proportional Stroke Recovery in the Rat." *Neurorehabilitation and Neural Repair* 32 (1): 3–6. https://doi.org/10.1177/1545968317751210.

Jeffers, Matthew Strider, Sudhir Karthikeyan, Mariana Gomez-Smith, Sarah Gasinzigwa, Jannis Achenbach, Astrid Feiten, and Dale Corbett. 2018. “Does Stroke Rehabilitation Really Matter? Part B: An Algorithm for Prescribing an Effective Intensity of Rehabilitation.” *Neurorehabilitation and Neural Repair* 32 (1): 73–83. https://doi.org/10.1177/1545968317753074.

Johansson, Barbro B., and Anna Lena Ohlsson. 1996. “Environment, Social Interaction, and Physical Activity as Determinants of Functional Outcome after Cerebral Infarction in the Rat.” *Experimental Neurology* 139 (2): 322–27. https://doi.org/10.1006/EXNR.1996.0106.

Kabra, Mayank, Alice A Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. 2013. “JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behavior.” *Nature Methods 2012 10:1* 10 (1): 64–67. https://doi.org/10.1038/nmeth.2281.

Kleim, Jeffrey A., Scott Barbay, Natalie R. Cooper, Theresa M. Hogg, Chelsea N. Reidel, Michael S. Remple, and Randolph J. Nudo. 2002. “Motor Learning-Dependent Synaptogenesis Is Localized to Functionally Reorganized Motor Cortex.” *Neurobiology of Learning and Memory* 77 (1): 63–77. https://doi.org/10.1006/nlme.2000.4004.

Kleim, Jeffrey A., and Theresa A. Jones. 2008. “Principles of Experience-Dependent Neural Plasticity: Implications for Rehabilitation After Brain Damage.” *Journal of Speech, Language, and Hearing Research* 51 (1): S225–39. https://doi.org/10.1044/1092-4388(2008/018).

Kleim, Jeffrey A, Jeffery A Boychuk, and Deanna L Adkins. 2007. “Rat Models of Upper Extremity Impairment in Stroke.” *ILAR Journal* 48 (4): 374–84.

Klein, Alexander, Lori-Ann R Sacrey, Ian Q Whishaw, and Stephen B Dunnett. 2012. “The Use of Rodent Skilled Reaching as a Translational Model for Investigating Brain Damage and Disease.” *Neuroscience and Biobehavioral Reviews* 36: 1030–42. https://doi.org/10.1016/j.neubiorev.2011.12.010.

Koch, C. E., B. Leinweber, B. C. Drengberg, C. Blaum, and H. Oster. 2017. “Interaction between Circadian Rhythms and Stress.” *Neurobiology of Stress* 6 (February): 57. https://doi.org/10.1016/J.YNSTR.2016.09.001.

Krakauer, John W., S. Thomas Carmichael, Dale Corbett, and George F. Wittenberg. 2012. “Getting Neurorehabilitation Right: What Can Be Learned from Animal Models?” *Neurorehabilitation and Neural Repair*. SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1177/1545968312440745.

Krakauer, John W., Asif A. Ghazanfar, Alex Gomez-Marin, Malcolm A. MacIver, and David Poeppel. 2017. “Neuroscience Needs Behavior: Correcting a Reductionist Bias.” *Neuron* 93 (3): 480–90. https://doi.org/10.1016/j.neuron.2016.12.041.

Krueger, Hans, Jacqueline Koot, Ruth E. Hall, Christina O’Callaghan, Mark Bayley, and Dale Corbett. 2015. “Prevalence of Individuals Experiencing the Effects of Stroke in Canada: Trends and Projections.” *Stroke* 46 (8): 2226–31. https://doi.org/10.1161/STROKEAHA.115.009616.

Kwakkel, Gert, Boudewijn Kollen, and Jos Twisk. 2006. “Impact of Time on Improvement of

Outcome After Stroke." *Stroke* 37 (9): 2348–53. https://doi.org/10.1161/01.STR.0000238594.91938.1E.

Labat-Gest, Vivien, and Simone Tomasi. 2013. "Photothrombotic Ischemia: A Minimally Invasive and Reproducible Photochemical Cortical Lesion Model for Mouse Stroke Studies." *J. Vis. Exp.* 76: e50370. https://doi.org/10.3791/50370.

Lai, Stefano, Alessandro Panarese, Cristina Spalletti, Claudia Alia, Alessio Ghionzoli, Matteo Caleo, and Silvestro Micera. 2015. "Quantitative Kinematic Characterization of Reaching Impairments in Mice After a Stroke." *Neurorehabilitation and Neural Repair* 29 (4): 382–92. https://doi.org/10.1177/1545968314545174.

Lang, Catherine E., Keith R. Lohse, and Rebecca L. Birkenmeier. 2015. "Dose and Timing in Neurorehabilitation: Prescribing Motor Therapy after Stroke." *Current Opinion in Neurology* 28 (6): 549–55. https://doi.org/10.1097/WCO.0000000000000256.

Lang, Catherine E., Jillian R. MacDonald, Darcy S. Reisman, Lara Boyd, Teresa Jacobson Kimberley, Sheila M. Schindler-Ivens, T. George Hornby, Sandy A. Ross, and Patricia L. Scheets. 2009. "Observation of Amounts of Movement Practice Provided During Stroke Rehabilitation." *Archives of Physical Medicine and Rehabilitation* 90 (October): 1692–98. https://doi.org/10.1016/j.apmr.2009.04.005.

Langhorne, Peter, Julie Bernhardt, and Gert Kwakkel. 2011. "Stroke Rehabilitation." *The Lancet* 377 (9778): 1693–1702. https://doi.org/10.1016/S0140-6736(11)60325-5.

Levitis, Daniel A, William Z Lidicker, and Glenn Freund. 2009. "Behavioural Biologists Do Not Agree on What Constitutes Behaviour." *Animal Behaviour* 78: 103–10. https://doi.org/10.1016/j.anbehav.2009.03.018.

Liao, Su Lan, Wen Yin Chen, Jon Son Kuo, and Chun Jung Chen. 2001. "Association of Serum Estrogen Level and Ischemic Neuroprotection in Female Rats." *Neuroscience Letters* 297 (3): 159–62. https://doi.org/10.1016/S0304-3940(00)01704-3.

Lieshout, Eline C. C. van, Julia Boonzaier, Adam J. Pel, Caroline L. van Heijningen, Jord J. Vink, Johanna M. A. Visser-Meily, Geralda A. F. van Tilborg, and Rick M. Dijkhuizen. 2021. "Translational Value of Skilled Reaching Assessment in Clinical and Preclinical Studies on Motor Recovery After Stroke:" *Neurorehabilitation and Neural Repair* 35 (5): 457–67. https://doi.org/10.1177/15459683211005022.

Lohse, Keith R., Catherine E. Lang, and Lara A. Boyd. 2014. "Is More Better? Using Metadata to Explore Dose-Response Relationships in Stroke Rehabilitation." *Stroke* 45 (7): 2053–58. https://doi.org/10.1161/STROKEAHA.114.004695.

Loong, Tze Wey. 2003. "Understanding Sensitivity and Specificity with the Right Side of the Brain." *British Medical Journal* 327 (7417): 716–19. https://doi.org/10.1136/bmj.327.7417.716.

MacLellan, Crystal L., Michael B. Keough, Shirley Granter-Button, Garry A. Chernenko, Stephanie Butt, and Dale Corbett. 2011. "A Critical Threshold of Rehabilitation Involving Brain-Derived Neurotrophic Factor Is Required for Poststroke Recovery." *Neurorehabilitation and Neural Repair* 25 (8): 740–48.

https://doi.org/10.1177/1545968311407517.

Macrae, IM. 2011. “Preclinical Stroke Research – Advantages and Disadvantages of the Most Common Rodent Models of Focal Ischaemia.” *British Journal of Pharmacology* 164 (4): 1062–78. https://doi.org/10.1111/J.1476-5381.2011.01398.X.

Mah, Kar Men, Abel Torres-Espín, Ben W Hallworth, John L Bixby, Vance P Lemmon, Karim Fouad, and Keith K Fenrich. 2021. “Automation of Training and Testing Motor and Related Tasks in Pre-Clinical Behavioural and Rehabilitative Neuroscience.” *Experimental Neurology* 340: 113647. https://doi.org/10.1016/j.expneurol.2021.113647.

Maldonado, Monica A, Rachel P Allred, Erik L Felthauser, and Theresa A Jones. 2008. “Motor Skill Training, but Not Voluntary Exercise, Improves Skilled Reaching After Unilateral Ischemic Lesions of the Sensorimotor Cortex in Rats.” *NeuroRehabilitation* 22: 250–61. https://doi.org/10.1177/1545968307308551.

Mathis, Alexander, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. 2018. “DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning.” *Nature Neuroscience* 21 (9): 1281–89. https://doi.org/10.1038/s41593-018-0209-y.

Mathis, Alexander, and Richard Warren. 2018. “On the Inference Speed and Video-Compression Robustness of DeepLabCut.” *BioRxiv*, 457242. https://doi.org/10.1101/457242.

McDonald, Matthew W., Kathryn S. Hayward, Ingrid C. M. Rosbergen, Matthew S. Jeffers, and Dale Corbett. 2018. “Is Environmental Enrichment Ready for Clinical Application in Human Post-Stroke Rehabilitation?” *Frontiers in Behavioral Neuroscience* 12 (July): 1–16. https://doi.org/10.3389/fnbeh.2018.00135.

Monfils, Marie-H, Erik J Plautz, and Jeffrey A Kleim. 2005. “In Search of the Motor Engram: Motor Map Plasticity as a Mechanism for Encoding Motor Experience.” *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry* 11 (5): 471–83. https://doi.org/10.1177/1073858405278015.

Montoya, C. P., L. J. Campbell-Hope, K. D. Pemberton, and S. B. Dunnett. 1991. “The ‘Staircase Test’: A Measure of Independent Forelimb Reaching and Grasping Abilities in Rats.” *Journal of Neuroscience Methods* 36 (2–3): 219–28. https://doi.org/10.1016/0165-0270(91)90048-5.

Moon, Seong Keun, Mariam Alaverdashvili, Albert R. Cross, and Ian Q. Whishaw. 2009. “Both Compensation and Recovery of Skilled Reaching Following Small Photothrombotic Stroke to Motor Cortex in the Rat.” *Experimental Neurology* 218 (1): 145–53. https://doi.org/10.1016/j.expneurol.2009.04.021.

Moro, Matteo, Giorgia Marchesi, Francesca Odone, and Maura Casadio. 2020. “Markerless Gait Analysis in Stroke Survivors Based on Computer Vision and Deep Learning: A Pilot Study.” *Proceedings of the ACM Symposium on Applied Computing*, March, 2097–2104. https://doi.org/10.1145/3341105.3373963.

Murphy, Timothy H., and Dale Corbett. 2009. “Plasticity during Stroke Recovery: From Synapse to Behaviour.” *Nature Reviews Neuroscience*. Nature Publishing Group.

https://doi.org/10.1038/nrn2735.

Nakayama, Hirofumi, Henrik Stig Jørgensen, Hans Otto Raaschou, and Tom Skyhøj Olsen. 1994. "Recovery of Upper Extremity Function in Stroke Patients: The Copenhagen Stroke Study." *Archives of Physical Medicine and Rehabilitation* 75 (4): 394–98. https://doi.org/10.1016/0003-9993(94)90161-9.

Nesin, Sibin Mathew, K. R. Sabitha, Anupam Gupta, and T. R. Laxmi. 2019. "Constraint Induced Movement Therapy as a Rehabilitative Strategy for Ischemic Stroke—Linking Neural Plasticity with Restoration of Skilled Movements." *Journal of Stroke and Cerebrovascular Diseases* 28 (6): 1640–53. https://doi.org/10.1016/J.JSTROKECEREBROVASDIS.2019.02.028.

Nica, Ioana, Marjolijn Deprez, Bart Nuttin, and Jean-Marie Aerts. 2018. "Automated Assessment of Endpoint and Kinematic Features of Skilled Reaching in Rats." *Frontiers in Behavioral Neuroscience* 11 (255). https://doi.org/10.3389/fnbeh.2017.00255.

Nudo, Randolph J. 2003. "Functional and Structural Plasticity in Motor Cortex: Implications for Stroke Recovery." *Physical Medicine and Rehabilitation Clinics of North America* 14 (1 SUPPL.): 57–76. https://doi.org/10.1016/S1047-9651(02)00054-2.

Nudo, Randolph J. 2013. "Recovery after Brain Injury: Mechanisms and Principles." *Frontiers in Human Neuroscience* 7: 887. https://doi.org/10.3389/fnhum.2013.00887.

Nudo, Randolph J., and Garrett W. Milliken. 1996. "Reorganization of Movement Representations in Primary Motor Cortex Following Focal Ischemic Infarcts in Adult Squirrel Monkeys." *Journal of Neurophysiology* 75 (5): 2144–49. https://doi.org/10.1152/jn.1996.75.5.2144.

Ohayon, Shay, Ofer Avni, Adam L Taylor, Pietro Perona, and S E Roian Egnor. 2013. "Automated Multi-Day Tracking of Marked Mice for the Analysis of Social Behaviour." *Journal of Neuroscience Methods* 219: 10–19. https://doi.org/10.1016/j.jneumeth.2013.05.013.

Peinemann, Alexander, Bibiana Reimer, Christian Löer, Angelo Quartarone, Alexander Münchau, Bastian Conrad, and Hartwig Roman Siebner. 2004. "Long-Lasting Increase in Corticospinal Excitability after 1800 Pulses of Subthreshold 5 Hz Repetitive TMS to the Primary Motor Cortex." *Clinical Neurophysiology* 115 (7): 1519–26. https://doi.org/10.1016/j.clinph.2004.02.005.

Pham, Jon, Sara M Cabrera, Carles Sanchis-Segura, and Marcelo A Wood. 2009. "Automated Scoring of Fear-Related Behavior Using EthoVision Software." *Journal of Neuroscience Methods* 178: 323–26. https://doi.org/10.1016/j.jneumeth.2008.12.021.

Plautz, Erik J, Garrett W Milliken, Randolph J Nudo, Scott Barbay, Kathleen Friel, Jeff Kleim, Diane Larson, and Haiying Wang. 2000. "Effects of Repetitive Motor Training on Movement Representations in Adult Squirrel Monkeys: Role of Use versus Learning We Thank Grey Gardner, Cami Knox, and Ramin Raiszadeh for Assistance with Data Collection, Patricia Pohl and Dennis Wallace for Helpful." *Neurobiology of Learning and Memory* 74: 27–55. https://doi.org/10.1006/nlme.1999.3934.

Remple, Michael S., Rochelle M. Bruneau, Penny M. VandenBerg, Crystal Goertzen, and Jeffrey A. Kleim. 2001. “Sensitivity of Cortical Movement Representations to Motor Experience: Evidence That Skill Learning but Not Strength Training Induces Cortical Reorganization.” *Behavioural Brain Research* 123 (2): 133–41. https://doi.org/10.1016/S0166-4328(01)00199-1.

Rioult-Pedotti, Mengia-S., Daniel Friedman, and John P. Donoghue. 2000. “Learning-Induced LTP in Neocortex.” *Science* 290 (October): 533–36. https://doi.org/10.1126/SCIENCE.290.5491.533.

Roman Siebner, Hartwig, and John Rothwell. 2003. “Transcranial Magnetic Stimulation: New Insights into Representational Cortical Plasticity.” *Exp Brain Res* 148: 1–16. https://doi.org/10.1007/s00221-002-1234-2.

Ruan, Jingsong, and Yao Yao. 2020. “Behavioral Tests in Rodent Models of Stroke.” *Brain Hemorrhages* 4 (1): 171–84. https://doi.org/10.1016/j.hest.2020.09.001.

Schaar, Krystal L, Miranda M Brenneman, and Sean I Savitz. 2010. “Functional Assessments in the Rodent Stroke Model.” *Experimental & Translational Stroke Medicine* 2 (13). https://doi.org/10.1186/2040-7378-2-13.

Schallert, T. 2006. “Behavioral Tests for Preclinical Intervention Assessment.” *NeuroRx* 3 (October): 497–504. https://doi.org/10.1016/J.NURX.2006.08.001.

Schallert, Timothy, Sheila M. Fleming, J. Leigh Leasure, Jennifer L. Tillerson, and Sondra T. Bland. 2000. “CNS Plasticity and Assessment of Forelimb Sensorimotor Outcome in Unilateral Rat Models of Stroke, Cortical Ablation, Parkinsonism and Spinal Cord Injury.” *Neuropharmacology* 39 (5): 777–87. https://doi.org/10.1016/S0028-3908(00)00005-8.

Scrivener, Katharine, Catherine Sherrington, Karl Schurr, and Daniel Treacy. 2011. “Many Participants in Inpatient Rehabilitation Can Quantify Their Exercise Dosage Accurately: An Observational Study.” *Journal of Physiotherapy* 57 (2): 117–22. https://doi.org/10.1016/S1836-9553(11)70022-4.

Shoyaib, Abdullah Al, Faisal F. Alamri, Abbie Biggers, Serob T. Karamyan, Thiruma V. Arumugam, Fakhrul Ahsan, Constantinos M. Mikelis, Taslim A. Al-Hilal, and Vardan T. Karamyan. 2021. “Delayed Exercise-Induced Upregulation of Angiogenic Proteins and Recovery of Motor Function after Photothrombotic Stroke in Mice.” *Neuroscience* 461 (May): 57–71. https://doi.org/10.1016/J.NEUROSCIENCE.2021.02.023.

Sturman, O, L von Ziegler, C Schläppi, F Akyol, M Privitera, D Slominski, C Grimm, et al. 2020. “Deep Learning-Based Behavioral Analysis Reaches Human Accuracy and Is Capable of Outperforming Commercial Solutions.” *Neuropsychopharmacology* 45 (11): 1942–52. https://doi.org/10.1038/S41386-020-0776-Y.

Tinbergen, Nikolaas. 1955. *The Study of Instinct*. Clarendon Press.

Truelsen, Thomas, Stephen Begg, and Colin Mathers. 2006. “The Global Burden of Cerebrovascular Disease.” 2006. https://www.who.int/healthinfo/statistics/bod_cerebrovasculardiseasestroke.pdf.

Ward, Nick S. 2004. “Functional Reorganization of the Cerebral Motor System after Stroke.”

*Current Opinion in Neurology* 17 (6): 725–30. https://doi.org/10.1097/00019052-200412000-00013.

Watson, Brant D., W. Dalton Dietrich, Raul Busto, Mitchell S. Wachtel, and Myron D. Ginsberg. 1985. "Induction of Reproducible Brain Infarction by Photochemically Initiated Thrombosis." *Annals of Neurology* 17 (5): 497–504. https://doi.org/10.1002/ana.410170513.

Weber, Rebecca Z, Geertje Mulders, Julia Kaiser, Christian Tackenberg, and Ruslan Rust. 2021. "Deep Learning Based Behavioral Profiling of Rodent Stroke Recovery." *BioRxiv*. https://doi.org/10.1101/2021.08.11.455647.

Whishaw, Ian Q., William T. O'connor, and Stephen B. Dunnett. 1986. "The Contributions of Motor Cortex, Nigrostriatal Dopamine and Caudate-Putamen to Skilled Forelimb Use in the Rat." *Brain* 109 (5): 805–43. https://doi.org/10.1093/brain/109.5.805.

Whishaw, Ian Q., Sergio M. Pellis, Boguslaw P. Gorny, and Vivien C. Pellis. 1991. "The Impairments in Reaching and the Movements of Compensation in Rats with Motor Cortex Lesions: An Endpoint, Videorecording, and Movement Notation Analysis." *Behavioural Brain Research* 42 (1): 77–91. https://doi.org/10.1016/S0166-4328(05)80042-7.

Whishaw, Ian Q., and Jo Anne Tomie. 1989. "Olfaction Directs Skilled Forelimb Reaching in the Rat." *Behavioural Brain Research* 32 (1): 11–21. https://doi.org/10.1016/S0166-4328(89)80067-1.

Whishaw, Ian Q, Forrest Haun, Bryan Kolb, and John Donne. 1999. "Analysis of Behavior in Laboratory Rodents." In *Modern Techniques in Neuroscience Research*, edited by U Windhorst and H Johansson, 1243–75. Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-58552-4_44.

Whishaw, Ian Q, and Sergio M Pellis. 1990. "The Structure of Skilled Forelimb Reaching in the Rat: A Proximally Driven Movement with a Single Distal Rotatory Component." *Behavioural Brain Research* 41: 49–59.

Zeiler, Steven R., and John W. Krakauer. 2013. "The Interaction between Training and Plasticity in the Poststroke Brain." *Current Opinion in Neurology* 26 (6): 609–16. https://doi.org/10.1097/WCO.0000000000000025.

Zhao, Shanshan, Mei Zhao, Ting Xiao, Jukka Jolkkonen, and Chuansheng Zhao. 2013. "Constraint-Induced Movement Therapy Overcomes the Intrinsic Axonal Growth–Inhibitory Signals in Stroke Rats." *Stroke* 44 (6): 1698–1705. https://doi.org/10.1161/STROKEAHA.111.000361.

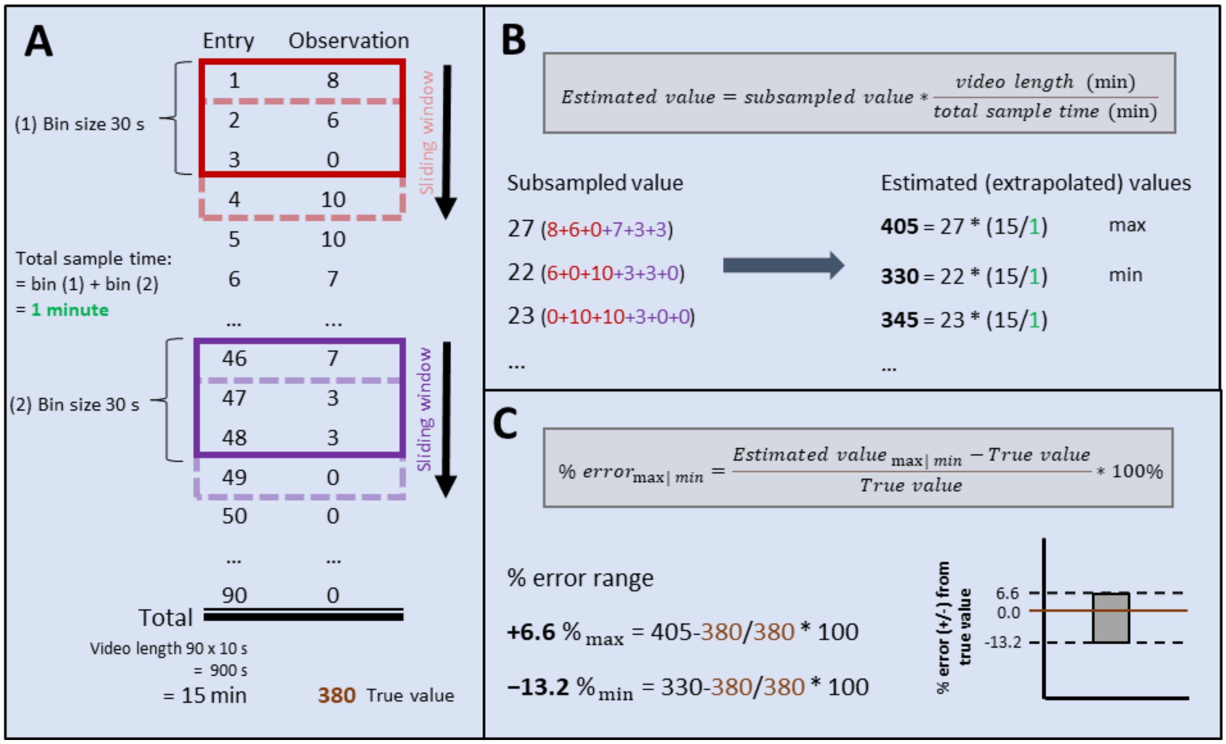


**Figure S 1. Sampling error analysis steps. A)** Rat reaching behavior in a full-length video (15 minutes) was denoted in table format whereby each entry represented seconds of "active reaching" in a 10-second period. A subsample of the data was created by using bins of various sizes (60, 30, 20, or 10 s) and varying the total number of bins used, such that the subsample contained either 1, 3, or 5 minutes of footage in total. The bins moved along the tabular data frame to create multiple subsamples of a video. **B)** The subsamples were extrapolated to give an estimated value of the number of reaches across the total video. To do this, the subsampled values were multiplied by a factor that would account for the fraction of footage that was used in the subsample. If 1 minute of data was used, it was multiplied by 15 to obtain an estimate for 15 minutes of video data. **C)** To determine the range of error (most positive to most negative), the maximum and minimum estimated values were used. Percent error was calculated as the % deviation of the estimated value from the true value. This range of error indicates the maximum deviation you can expect when subsampling using a particular bin size and total sample time combination.

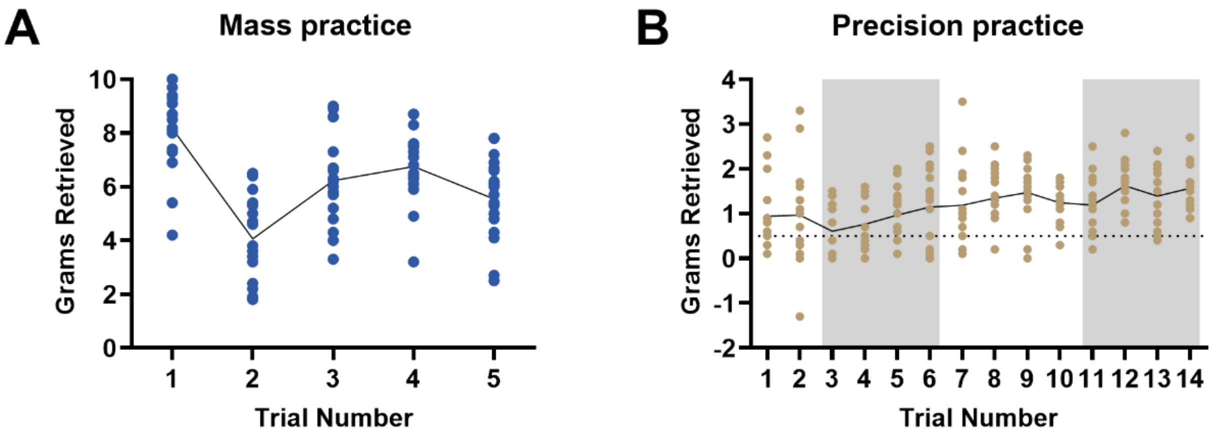


**Figure S 2. Pre-stroke reach training data. A)** Mass practice training occurred for 3 days (n = 19). In trial 1 the trough was placed at mid-height, enabling easy access to the pellets. In trial 2 to 5, the trough was placed at a low height. **B)** Precision practice training occurred over 4 days (n = 20), which are indicated by alternating white and grey shaded areas. Trial 1 and 2 consisted of manual pellet presentation on a tray. From trial 3 onwards the automated pellet presenter device was used. We had 15 precision practice units available for rehabilitation, and from pilot cohorts, we learned that approximately 25% of the rats do not fully engage in the task and will grasp on average <0.5 g (11 pellets). Therefore, this was used as a criterion to be included in the rehabilitation paradigm, indicated by the dashed line. The black lines in each graph indicate the group average.

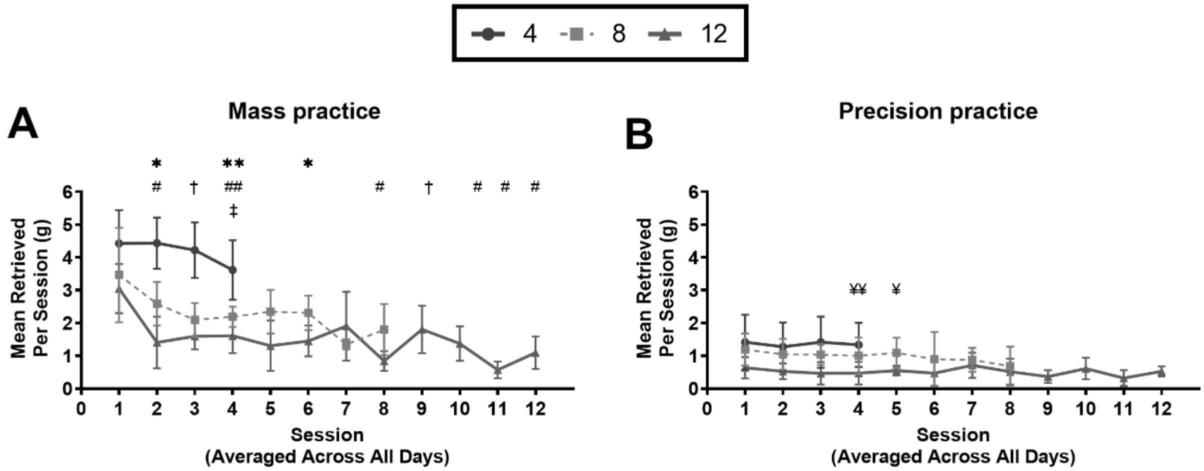


**Figure S 3. Progression of average pellet retrieval per session across rehabilitation day. A)** Mass practice (n = 6). Animals with 4 sessions showed no main effect of session number on mean pellets retrieved one-way ANOVA, F (2.409, 12.04) = 1.630, *p* = .2361. Average pellet retrieval declined 0.8 g between session 1 (4.4 ± 1.02) and 4 (3.6 ± 0.91). In the group with 8 sessions, there was a main effect of session number on pellet retrieval: one-way ANOVA, F (2.299, 11.50) = 5.850, *p* = .0152. Tukey's post-hoc analysis revealed there was a downward trend from session 1 to 2 (-0.9 g, *p* = .8204) and from 2 to 3 (0.5 g, *p* = .2868) after which performance was relatively steady between until session 6. Session 7 on average had the lowest pellet retrieval (1.3 g ± 0.15), which was significantly less compared with session 2 (2.6 g ± 0.67, *p* = .0480), session 4 (2.2 g ± 0.31, *p* = .0057), and session 6 (2.3 ± 0.53, *p* = .0430). Main effect of session number was also revealed in the group with 12 sessions per day (one-way ANOVA, F (2.803, 14.02) = 7.158; *p* = .0042. There was an immediate decline in pellets retrieved after the first training session. In session 1 rats retrieved 3.1 g ± 0.75 pellets. This was more than session 2 (1.4 g ± 0.79, *p* = .0463), session 4 (1.6 g ± 0.53, *p* = .0062), session 8 (0.8 g ± 0.31, *p* = .0114), session 10, 11, and 12 (1.4 g ± 0.52, *p* = .0256; 0.6 g ± 0.26, *p* = .0135; 1.1 g ± 0.50, *p* = .0004, respectively). Furthermore, in session 3 and 9 more pellets were retrieved compared to session 11 (1.6 g ± 0.40, *p* = .0385; 1.8 g ± 0.31, *p* = .0493). And more pellets were retrieved in session 4 vs. 12 (*p* = .0385). **B)** Precision practice (n = 5). A one-way repeated measures ANOVA did not demonstrate a main effect of session number on pellet retrieval in the 4 sessions group (F (2.409, 12.04) = 1.630, *p* = .2361). In the group with 8 sessions there was no main effect of session number on pellet retrieval F (1.994, 7.975) = 2.676, *p* = .1291. However, Tukey's post-hoc analysis revealed a significant decline in the second half of the day, between session 4 (1.0 g ± 0.56) and 5 (1.1 g ± 0.46) vs. 8 (0.7 g ± 0.60), (*p* = .0028; *p* = .0369, respectively). In the group with 12 sessions average pellet retrieval was constant (one-way ANOVA, F (3.015, 12.06) = 2.043, *p* = .1612). * *p* < .05, ** *p* <0.01, difference from day 7, 8 session group; # *p* < .05, ## *p* < .01, difference from day 1, 12 session group; † *p* < .05 difference from day 11, 12 session group; ‡ *p* < .05 difference from day 12, 12 session group; ¥ *p* < .05, ¥¥ *p* < .01, difference from day 8, 8 session group precision practice. Data are mean ± SD.

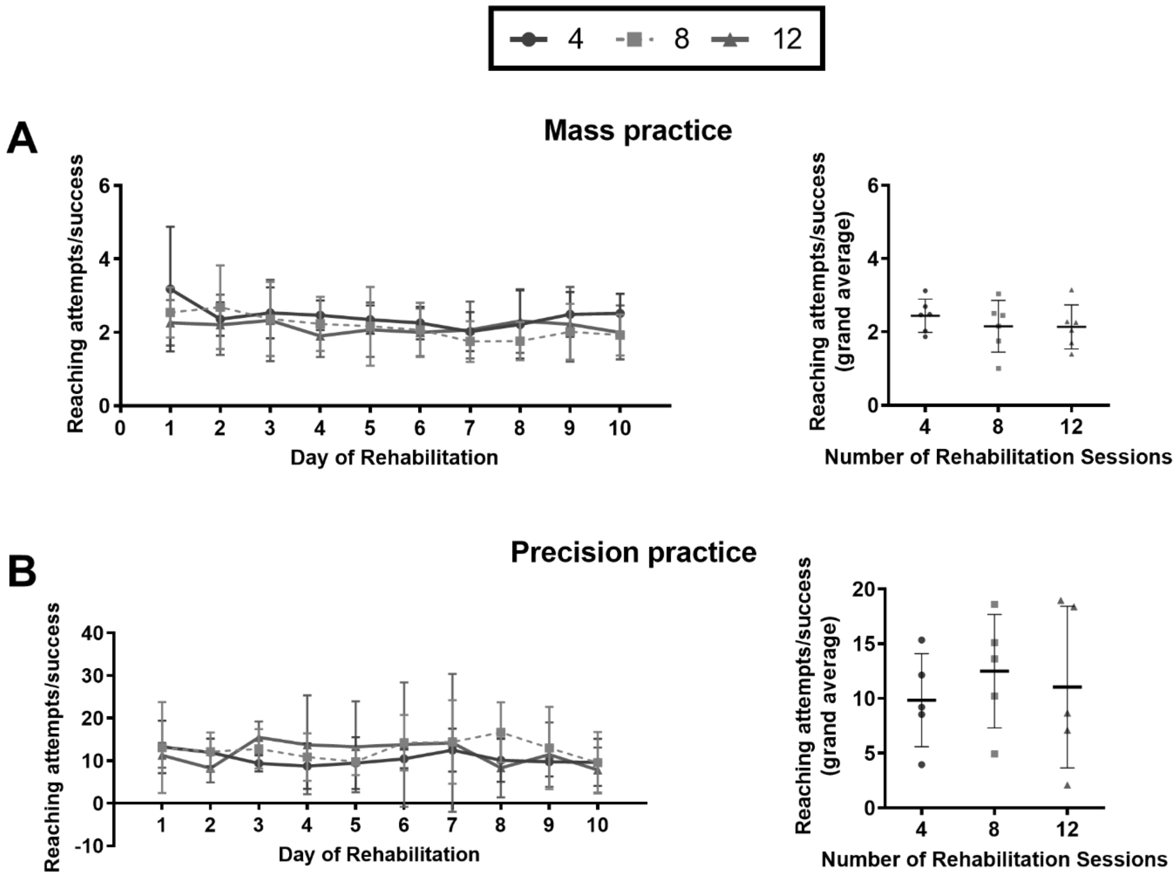


**Figure S 4. Retrieval efficiency.** Retrieval efficiency represents the ratio between the number of reaching attempts and pellet retrieval. The left graphs show mean retrieval efficiency across rehabilitation and the right graphs show the grand average, for both **(A)** mass (n = 6) and **(B)** precision practice (n = 5). Rats in mass practice perform ~2.2 reaches per pellet retrieval and in precision practice ~11.06. Although reaching attempts are more ambiguous in the precision paradigm (due to hovering movements), it shows that rats put in more effort in this task to obtain a pellet. Data are mean ±SD.

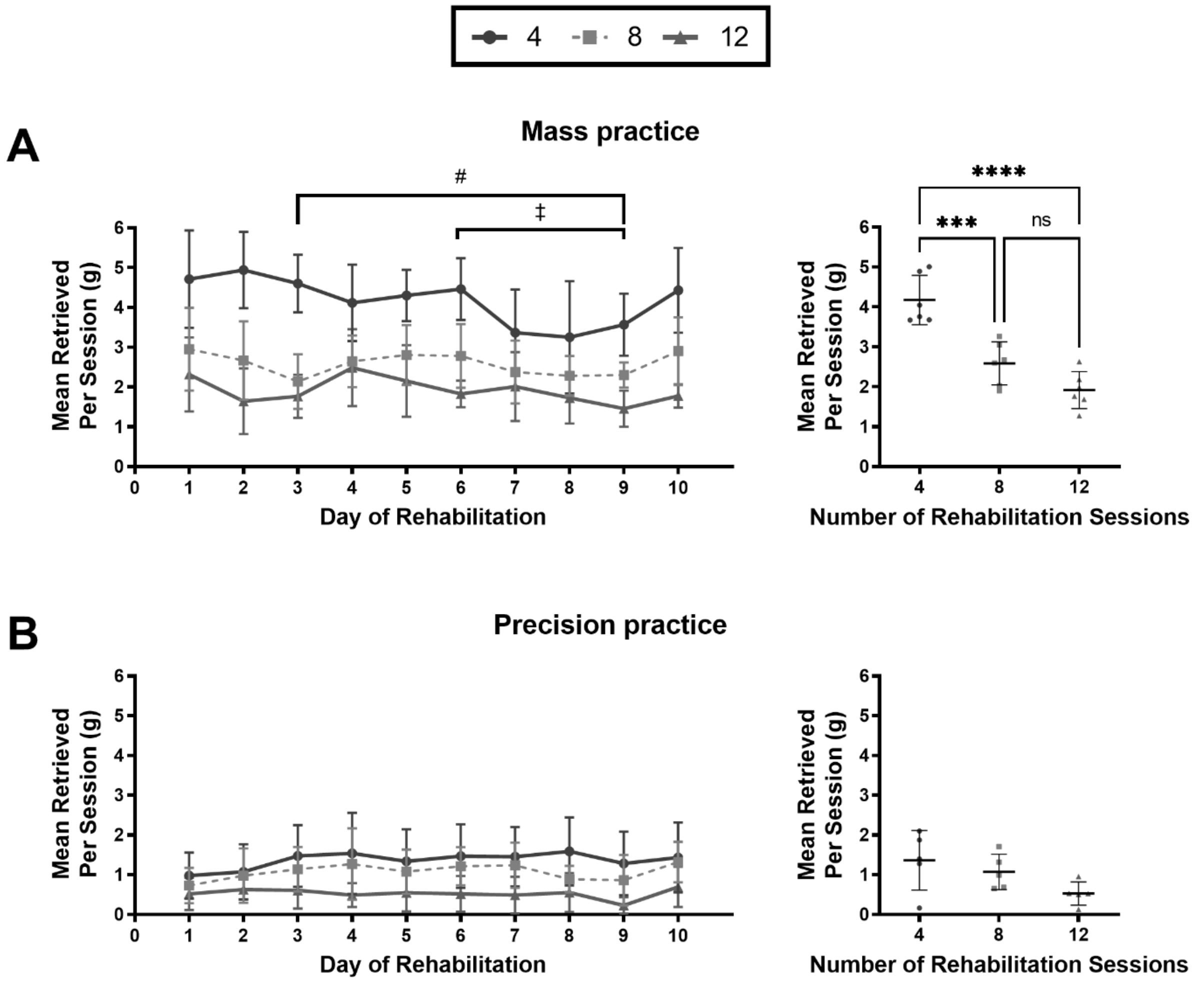


**Figure S 5. Mean grams retrieved in first for 4 sessions per day.** The average grams retrieved were calculated using only the first 4 daily sessions of each rehabilitation group. This enables a direct comparison of the effect of resting time between sessions on training performance. **A)** In mass practice (n = 6) there was a significant interaction effect of time x group (F(18, 135) = 1.966, *p* = .0156), a main effect of day (F(2.559. 38.38) = 4.625, *p* = .0102), and a main effect of group (F(2, 15) = 27.24, *p* < .0001) on pellet retrieval (two-way ANOVA with Tukey's post-hoc comparisons). A one-way ANOVA with Tukey's post-hoc comparisons revealed that animals with 4 sessions (4.175 g ± 0.62) still grasp significantly more pellets compared with rats with 8 (2.585 g ± 0.54, *p* = .0004), and 12 sessions (1.917 g ± 0.46, *p* < .0001). This is in line with the results shown in **Figure S 3.** Although the performance gap is small during session 1, performance quickly declines in the groups with 8 and 12 sessions. This decrease could be attributed to insufficient recovery time between sessions. **B)** In the precision practice (n = 5), there was no interaction effect, nor a main effect of time or group on pellet retrieval (two-way ANOVA). Moreover, overall group differences were non-significant (one-way ANOVA). These data imply that resting time does not have a significant impact on performance in the precision reaching task. # *p* < .05, post-hoc effect in group with 4 sessions, ‡ *p* < .05 indicates post-hoc effect in group with 12 sessions, *** *p* < .001, **** *p* < .0001, ns = non-significant. Data are mean ± SD.

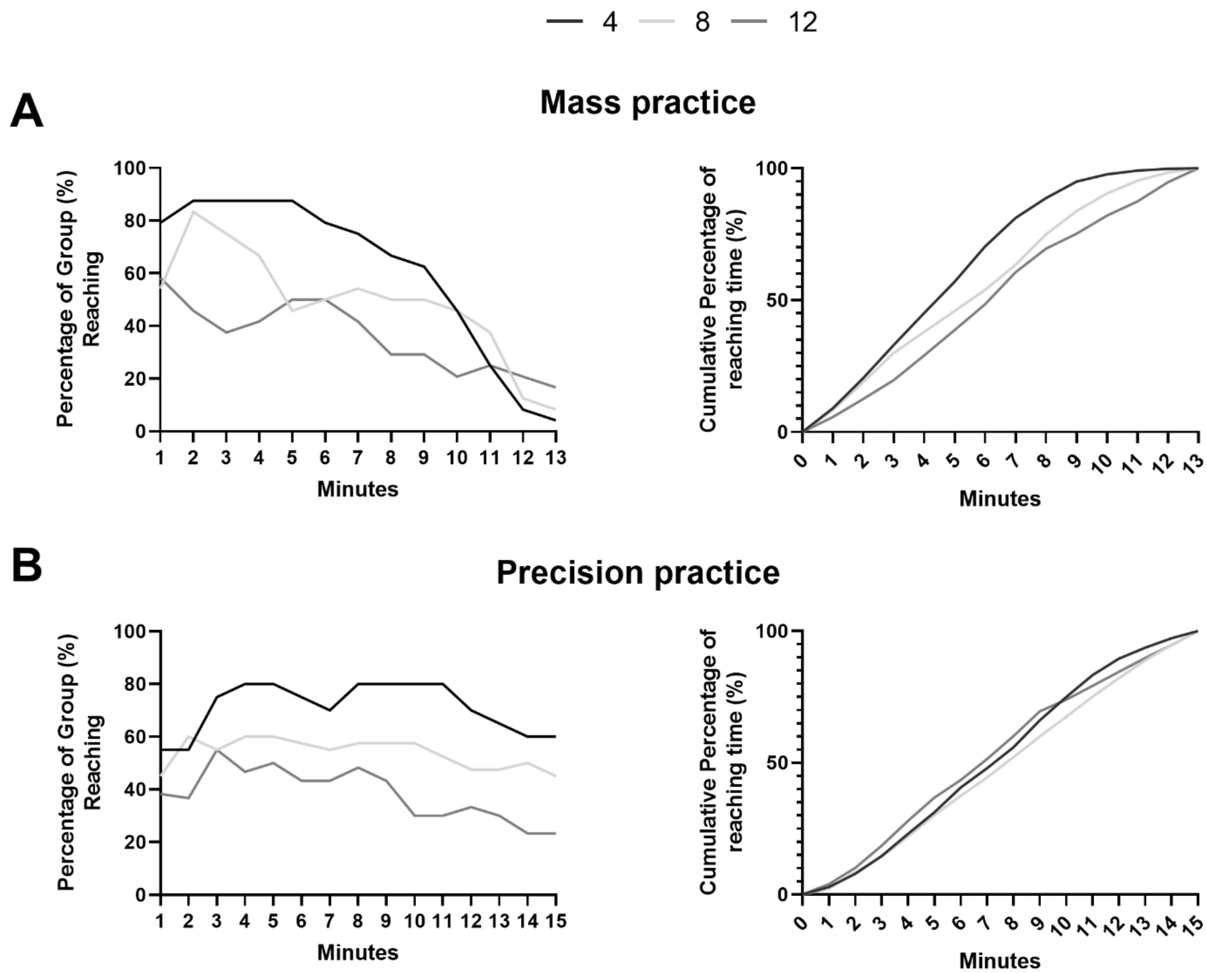


**Figure S 6. Within session reaching activity distribution per minute.** Reaching time, defined as the amount of time the paw was located within the region of interest with the reach target, was used to describe the rats' reaching activity. This was monitored per minute of video footage. A cut-off of >0.33 s / minute was set to consider a rat being engaged with the target for that minute. Data are based on one rehabilitation day. **Left:** shows the average percentage of each training group that demonstrated activity per minute of a training session for both **(A)** mass (n = 6) and **(B)** precision practice (n = 5). In mass practice, there was a strong decline in the number of animals that remain engaged towards the end of the session. In precision practice, rats demonstrate more continuous activity. In both paradigms, the group with 4 sessions had the most rats reaching for the majority of the training session. **Right:** shows the cumulative percentage of reaching time. In mass practice, rats with 4 sessions will perform 95% of their activity during the first 9 minutes of the session. Rats with 8 and 12 sessions reach this point at 11 and 12 minutes, respectively. This is in line with the steep decline that occurs after the ninth minute for the group with 4 sessions in the left graph, and a subsequent decline of the other groups. In precision practice, all groups achieve 95% of their reaching time after approximately 14 minutes. This corresponds to the more continuous reaching activity observed in the left graph.