

Inverse Optimal Control Model of Driving Behavior in Depressed Individuals

He Crane Huang (heh001@ucsd.edu)

Department of Cognitive Science, UC San Diego
9500 Gilman Drive, La Jolla, CA 92037, USA

Katia M. Harlé (katia.harle@gmail.com)

Department of Psychiatry, UC San Diego
9500 Gilman Drive, La Jolla, CA 92037, USA

Martin Paulus (mpaulus@ucsd.edu)

Department of Psychiatry, UC San Diego
9500 Gilman Drive, La Jolla, CA 92037, USA

Javier Movellan (jrmovellan@gmail.com)

Machine Perception Lab, UC San Diego, Emotient
La Jolla, CA 92037, USA

Abstract

Poor performance in goal-oriented sensory motor tasks is a common symptom among depressed individuals. However, it is unclear what the underlying causes of these deficits are. Elucidating the underlying mechanisms is an important first step to develop more targeted behavioral interventions. Here, using simple motor-control tasks, we propose an inverse optimal control approach to analyze and factorize performance deficits into two components of subjects' behaviors: 1) sensory motor speed, 2) reward-processing. In Task 1, subjects with Beck Depression Inventory score ranging from 0 to 36 were instructed to push a joystick as quickly as possible once they observe motion onset of a virtual car. In Task 2, they were instructed to drive a virtual car as quickly as possible and stop it as close as possible to a stop sign. Based on the continuous joystick actions for each individual subject, we estimated perceptual motor efficiency parameters and recovered the underlying reward function that best explained the subject's behavior. Initial results suggest, that relative to healthy controls, depressed individuals: 1) have deficits in sensory-motor processing speed, 2) have different goals but not significantly different accuracy/effort ratio. The results suggest that inverse optimal control may be a viable computational approach to quantify and factorize the underlying causes of sensory motor deficits in individuals with depression.

Keywords: depression; motor-control; computational model; inverse optimal control; inverse reinforcement learning; reward-processing.

Introduction

Depression can affect many facets of daily life. It accounts for 8.2% of global years lived with a disability (YLDs) in 2010, and has become a worldwide health priority (World Healthy Organization, 2012). In particular, growing evidences show depression increases the odds ratio for a car accident (Chapman & Perry, 2008), and reduces driving performance in a driving simulator (Wingen et al. 2006). For instance, Selzer et al. (1968) reported for fatal driving accidents, 21% of drivers were clinically depressed, compared with 7% in healthy controls. Hilton et al. (2009) reported severe and very severe depression was associated with an increased odds ratio for being involved in an accident or near miss in the past 28 days. In a driving simulator, Bulmash et

al. (2006) found depressed individuals exhibited slower reaction times and increased number of crashes when compared to controls. Despite compelling evidence of the severe consequences from poor motor-control in depressed individuals, the influence of depressive mood on driving actions remains largely unknown. However, so far, there are few studies (clinical or basic) have emphasized this issue.

Driving task is a closed-loop feedback control process (Lenard Evans, 2004). Drivers assess current driving environment from sensory feedback, and make control commands based on the goal. The decisions a driver makes given a task are in a hierarchical system (Janssen 1979) that comprises 1) a strategic level that is associated with one's goal in driving (motivation) and 2) a control level that is associated with one's sensorimotor skills (perceiving sensory feedback and executing motor commands). Thus depressive symptoms that influence those two levels can lead to different driving behavior.

Motivation deficits in depression It has been shown that depressed individuals have greater sensitivity to risk and punishment (Trew 2011), while risk is the most common of all the motivations considered by driving researchers. Risk homeostatis theory (Taylor, 1964) postulated that drivers adjust their speed in accordance with the perceived risk. However, risk perception differs greatly among individuals. It is affected not only by the objective danger in the situation (weather, road condition), but also by the driver's own assessment of his or her actions (e.g., driving faster than legal speed limit). Thus for a driving task that is considered almost as risk-less to an experienced F1 driver, it may be perceived as highly risky to depressed individuals.

Recently, Wilde (2002) proposed that we drive not to minimize risk (or maximize safety), but to reduce or increase it to a desired risk level with which we feel comfortable. The target level of risk varies among drivers. For young drivers

who are more risk seeking than others, they may set a higher level of risk to fulfill the thrill of driving. However, for depressed individuals who are oversensitive to punishment, they probably have a different driving style that targets a lower risk level. In other words, they may have differences in reward-processing that favors goals with lower risk. In this study, we will investigate what are the differences in their reward-processing, how that influences their driving actions, and what it implies about their risk sensitivity.

Psychomotor disturbance On the other hand, depressed individuals suffer from psychomotor disturbance (Buyukdura et al. 2011). Behavioral experiments have suggested impairments in sensorimotor system among depressed individuals. Caligiuri & Ellwanger (2000) showed that depressed individuals have difficulties performing normal physical actions, such as simple motor learning tasks. Sabbe et al. (1999) used simple drawing tasks and showed that MDD (Major Depressive Disorder) patients exhibited marked motor deficits of the visuomotor control process (longer movement duration, longer pauses, and lower velocities). Those sensorimotor impairments will adversely affect how one responds to sensory feedback and executes motor commands in driving tasks.

Thus, poor motor-performance may be a consequence of mixed depressive symptoms. It could be due to 1) different targeted risk level (goals), and/or 2) impaired sensorimotor system. Our study aims to provide a computational approach (inverse reinforcement learning) to disentangle these processes.

Inverse Optimal Control

In optimal control theory (inverse reinforcement learning at continuous time), actions are chosen to optimize a performance criterion (Todorov & Jordan 2002). The performance criterion is defined as a reward-function that includes task-related performance measure and action cost. For example, in a task that instructs subjects to drive to a location A as quickly as possible, the performance measure can be the stopping distance to A, and the action cost can be the accumulated effort of accelerating and decelerating controls. Different individuals may have different target stopping distance to A, and different weights to assess the ratio of the closeness to the target location over the action cost (i.e. accuracy/effort ratio), thereby forming different reward-functions.

With different reward-functions in mind, there will be different action-planning strategies, which are defined as control-policies. A control-policy comprises a series of dynamic decisions modulating actions at given states in continuous time (Shadmehr 2008). In a forward model, with experimentally defined reward function (for example, points), we can derive the optimal control-policy to optimize the reward function. In an inverse model (Ng & Russell, 2000), with observed continuous actions, we can infer the control-policy, and recover the reward-function used in developing this control-policy. Thus the objective of inverse optimal control is to infer individuals' reward-function based on observed

behaviors. This approach will provide a quantitative comparison of how different reward-processing between depressed and healthy controls lead to observed behavioral differences.

In summary, we will apply inverse reinforcement learning approach to investigate how reward-processing and sensorimotor impairments in depressed individuals influence their motor control in a simulated driving task.

Method

Participants

58 college students (15 male and 43 female subjects) in UCSD participated this study in fall quarter 2013. They signed up through UCSD SONA system, and then completed phone-screening and on-line BDI (Beck Depression Inventory, BDI-II, Beck et al. 1996) measure. Qualified subjects completed the experiment (with a second BDI measure prior to the task) in the lab, and were compensated by 2 course credits. Their onsite BDI range from 0 to 36 with mean BDI=10.25 (std=8.38), median BDI=8.

Experiment

Subjects were instructed to complete two tasks in this experiment. Both tasks were computer experiments (on a 15 inch MacBook Pro) programmed in Matlab. We recorded their continuous actions using a gaming joystick (Thrustmaster HOTAS Warthog Flight Stick). The goal of Task 1 (Move-and-Go) is to measure individual's perceptual and motor speed (without risk influence), and the goal of Task 2 (Speed-and-Stop) is to apply inverse optimal control model to recover reward-function (with risk influence).

Task 1: Move-and-Go Subjects were required to perform Task 1 twice (120 trials, before and after Task 2). In each trial (Figure 1), a car would appear on the bottom of the screen, and subjects were instructed to push the joystick from resting position forward to the maximum position as quickly as possible once they observe the car move. Each trial started with a 3-second countdown and a random waiting interval (1-3 seconds), then the car would start to move at a randomly selected speed (.01-.3 cm/second). Trials ended once subjects pushed the joystick at its maximum forward position. The goal in this task is risk-free, thus parameters estimated here can be considered to represent basic sensorimotor skills.

Task 2: Speed-and-Stop There were 3 blocks, with 20 trials/block in Task 2. In each trial (Figure 2), subjects were instructed to drive a virtual car as quickly as possible to a stop sign (distance: 10.62 cm) without crossing the stop-line, and stop there within a 10-second time window. Each trial started with a 3-second countdown and ended when time ran out, with no performance feedback (e.g., points) in the end. The car has a linear dynamic system (see Model), in which the car position is controlled by continuous joystick position.

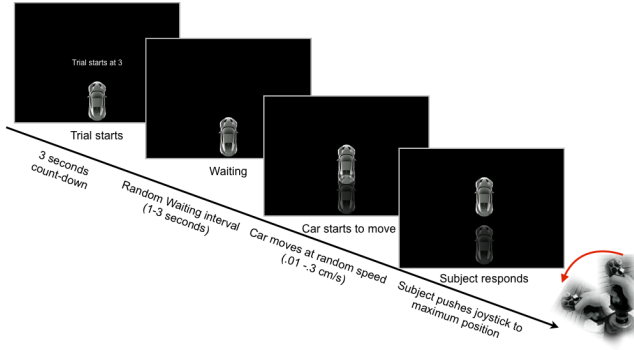


Figure 1: Task 1 (Move-and-Go)

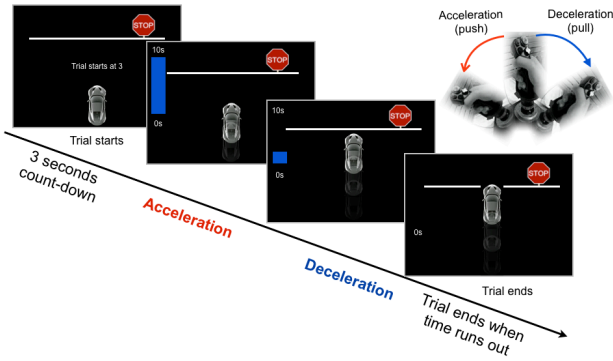
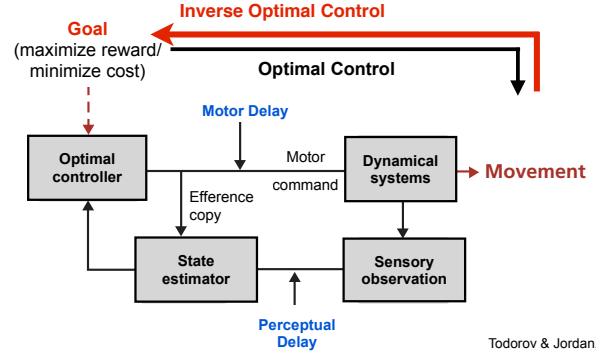


Figure 2: Task 2 (Speed-and-Stop)

Model

Driving task is a dynamic process of sensorimotor integration (Flanders, 2011), in which the brain (optimal controller) takes sensory information and uses it to make continuous motor actions. In this process (Figure 3), the optimal controller estimates the current state at time t , produces a motor command based on the goal and keeps an efference copy (the expected outcome of the motor command) at the state estimator, and sends the motor command to muscles to generate the movement. Then the state estimator will update the efference copy with the delayed sensory observation to predict state at next time point $t + 1$ and the optimal controller will generate new motor commands until the goal is reached.

We propose to use inverse optimal control model to explain observed behavior in this feedback control process. To achieve that, we first assessed individual's sensorimotor system by estimating their perceptual speed (delay in perceiving sensory observation at time t) and motor speed (delay in executing motor command at time t) in a risk-free task (Task 1: move-and-go). Then we estimated their target state (target stopping distance) and target accuracy/effort ratio (the willingness to reach the target state) in the reward function in Task 2 (speed-and-stop) with the perceptual and motor delay parameters from Task 1.



Todorov & Jordan, 2002
Stephen H. Scott 2002

Figure 3: Model Framework

Perceptual speed γ and motor speed β

Task 1 (move-and-go) was designed to estimate perceptual speed γ and motor speed β . We model subjects' perceived car position Y_t as a delayed true car position X_t due to the limit of sensory processing speed γ (Eq.1). The higher the γ , the closer the perceived car position Y_t to the true car position X_t . We assume subjects will decide the car starts moving once the perceived car position Y_t reaches a position threshold X_{thd} . Thus the minimal time for the perceived car position Y_t to reach the threshold X_{thd} is reaction time t_{RT} (Equation 2):

$$\text{Perceived car position } Y_t : dY_t = \gamma(X_t - Y_t)dt \quad (1)$$

$$\text{Reaction Time : } t_{RT} = \underset{t}{\operatorname{argmin}}\{Y_t \geq X_{thd}\} \quad (2)$$

We model joystick position C_t as a delayed execution from target joystick position U_{target} , due to the limit of motor execution speed β (Equation 3). The higher the β , the closer joystick action to the desired target position. Thus the minimal time for C_t to reach U_{target} is movement time (Equation 4).

$$\text{Joystick position } C_t : dC_t = \beta(U_{target} - C_t)dt \quad (3)$$

$$\text{Movement Time : } t_{MT} = \underset{t}{\operatorname{argmin}}\{C_t \geq U_{target}\} \quad (4)$$

In above equations, X_t (true car position), t_{RT} (reaction time to car motion-onset), C_t (recorded joystick position), U_{target} (target position) and t_{MT} (movement time) are known. We use t_{RT} and X_t to recover X_{thd} , γ and Y_t , and use C_t and t_{MT} to recover β , by optimizing over γ , X_{thd} , and β to give the minimal errors between predicted t_{RT} , t_{MT} and observed data.

Inverse optimal control of the driving task

Task 2 (speed-and-stop) was designed to estimate individual's reward-function, which is a function of target stopping distance and accuracy/effort ratio. Target stopping distance measures individual's risk sensitivity. The further away one aims to stop from the stop sign, the less risk there is to cross the stop-line. Target accuracy/effort ratio measures individual's willingness to reach the target stopping distance. The higher

the ratio, the more motivated one is to stop as close as possible to the target stopping location. In a quadratic reward function, target distance represents the optimal point of the reward function, and target accuracy/effort ratio represents the hessian of the reward function.

Linear Quadratic Gaussian Model (LQG) We formulate the driving task as a LQG problem with a linear dynamic system and a quadratic reward function. In forward LQG problems, the optimal controller generates an optimal control policy that maximizes a given reward function. Figure 4 shows in a forward model of this driving task, how different model parameters (motor speed β , target accuracy/effort ratio P , and target stopping distance X_{target}) can affect optimal car position and joystick control. In inverse LQG problems, we use observed movements to infer the underlying reward function that best explains the observed behavior.

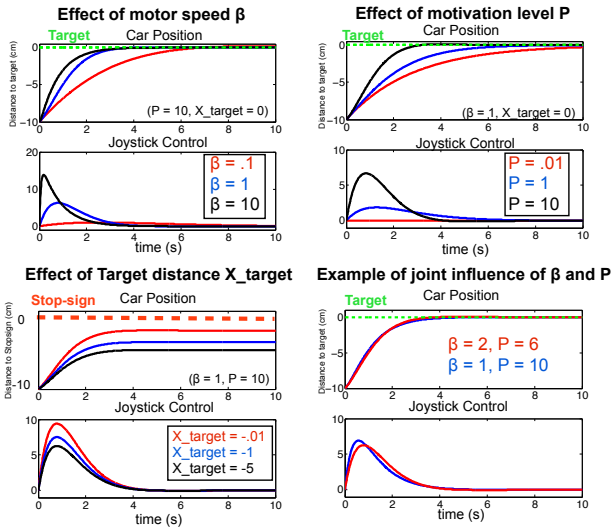


Figure 4: Influences of model parameters. β : higher motor speed lead to faster arrival time to target; P : higher motivational level lead to faster arrival time and closer distance to target; X_{target} : different target distances lead to different stopping position; Joint influence of β and P : similar behavior may have very different underlying causes. Someone with higher motor speed ($\beta = 2$) and lower accuracy/effort ratio ($P = 6$) may have similar behavior as someone with lower motor speed ($\beta = 1$) but higher accuracy/effort ratio ($P = 10$).

Linear dynamic system Assuming the driving task as a linear dynamic system (Equation 5) with a partial hidden state X_t and observable feedback Z_t , in which X_t is a 3x1 vector including the (hidden) true car distance to target stopping position at time t , joystick action at time t , and perceived car distance to target stopping position at time t .

Partial observable linear system: $dX_t = AX_t dt + BU_t dt$ (5)

Observation: $Z_t = CX_t + V_t$ (6)

With:

$$A = \begin{bmatrix} a & b & 0 \\ 0 & -\beta & 0 \\ \gamma & 0 & -\gamma \end{bmatrix} \quad (7)$$

$$B = \begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix} \quad (8)$$

$$C = [0, 0, 1] \quad (9)$$

In which, a, b are car dynamics parameters (assuming known), V_t is Gaussian noise, β and γ are motor and perceptual speed that are estimated from Task 1. Note that in the state X_t , the hidden true car position and perceived car position are measured as a distance to target stopping position (parametrized as the target state in the reward function), which we will estimate through optimization from this model.

Quadratic reward function We assume the reward function $r(X_t, U_t)$ is a function that evaluates the state X_t (through $g(X_t)$) and the action U_t (through $U_t^2 q$).

$$\text{reward function: } r(X_t, U_t) = g(X_t) - U_t^2 q \quad (10)$$

Without loss of generality, let $q = 1$ (i.e. optimal action will not change if scaling the reward function), thus $g(X_t)$ is a function of target state and target accuracy/effort ratio. We assume subjects were using a stationary (infinite horizon) policy and the reward function has a diagonal form (i.e. no joint influence between state elements in the reward function).

In LQG setting, subjects first estimate true state from observation using a Kalman filter to convert the problem to a fully observable system, and then solve it as a LQR (Linear-Quadratic-Regulator) problem:

$$d\hat{X}_t = A\hat{X}_t dt + BU_t dt + L_t(Z_t - C\hat{X}_t)dt \quad (11)$$

$$U_t = -K\hat{X}_t \quad (12)$$

In which L_t is Kalman gain. U_t is a linear combination of the states and K can be estimated from U_t and recorded behavior data through linear regression. This suggests a quadratic value function:

$$v(\hat{x}, t) = -\frac{1}{2}\hat{x}'_t w \hat{x}_t \quad (13)$$

Then the HJB equation (Bellman, 1957) for this linear system will give us $g(\hat{x})$ as a quadratic form of \hat{x} :

$$g(\hat{x}) = -\frac{1}{2}\hat{x}'(-2A'w + k'k)\hat{x} \quad (14)$$

In which we define P as the target accuracy/effort ratio:

$$g(\hat{x}) = -\frac{1}{2}\hat{x}'P\hat{x} \quad (15)$$

$$P = -2A'w + k'k \quad (16)$$

In which A and k are known from equation (7) and (12), and w can be solved by using optimal LQR solution.

Results

Task 1: Move-and-Go

The purpose of this task was to estimate perceptual and motor speed for individual subject, and use those estimation in the inverse optimal control model.

Perceptual-motor speed Figure 5A (scatterplot) shows as BDI increases, reaction time and movement time increases, which suggests slower perceptual and motor speed in depressed individuals. Our model results are consistent with observed behavior (Figure 5B).

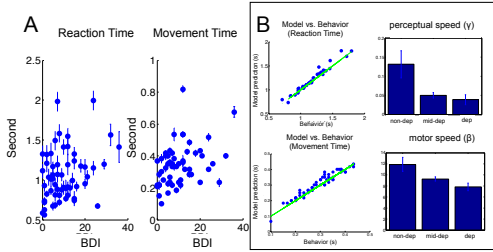


Figure 5: A: Reaction time to car motion onset and movement time to push joystick to maximum forward position; B: Left: model vs. data, points represent individual subjects. Right: group comparison of perceptual and motor speed. (non-dep: $BDI \leq 5$, mid-dep: $6 \leq BDI < 20$, dep: $BDI \geq 20$)

Task2: Speed-and-Stop

The purpose of this task was to estimate the reward function that best explained each subject's behavior, taking account of individual's perceptual-motor speed estimated from Task 1. The reward function consists of two components: 1) target stopping distance from stop sign and 2) target accuracy/effort ratio.

Target stopping distance By categorizing subjects into 3 groups based on their BDI (non-dep: $BDI \leq 5$, mid-dep: $6 \leq BDI < 20$, dep: $BDI \geq 20$), Figure 6 (left) shows the differences in their stopping distance over time: 1) non-dep group has the closest target distance while dep group has the furthest target distance to stop sign; 2) Non-dep and mid-dep group have relatively stable target distances throughout the experiment, but dep group has a continuously increasing stopping distances with increasing variability over time. Target stopping distance estimated from the inverse model (Figure 6 Right) are consistent with above behavioral result. Examples of stopping position overtime from non-dep and dep group are shown in Figure 7.

Target accuracy/effort ratio Taking account of different target stopping distances (Figure 6) in reward-processing, model results (Figure 8A) shows the mean of accuracy/effort ratio in depressed group is not significant different from healthy controls. Examples of model prediction in continuous time are shown in Figure 8B.

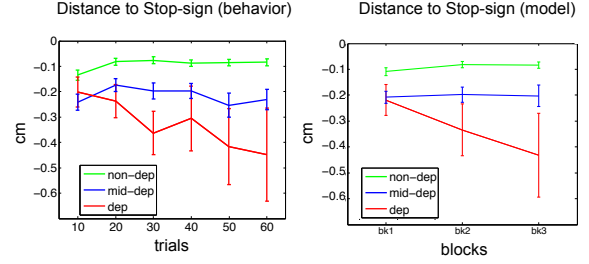


Figure 6: Left: Observed stopping distance over time (60 trials). Right: Model prediction of targeted stopping distance over blocks (3 block, 20 trial/block).

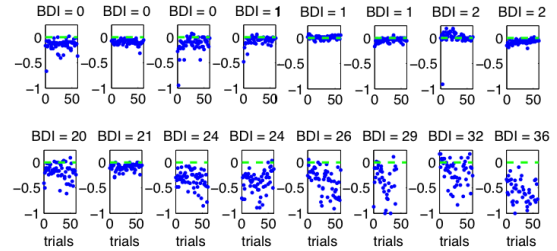


Figure 7: Examples of individual stopping distance over time. Top row: examples from 8 healthy controls. Bottom row: examples from 8 depressed individuals. X-axis: trials; Y-axis: stopping distance to target (cm); green dash line: stop sign.

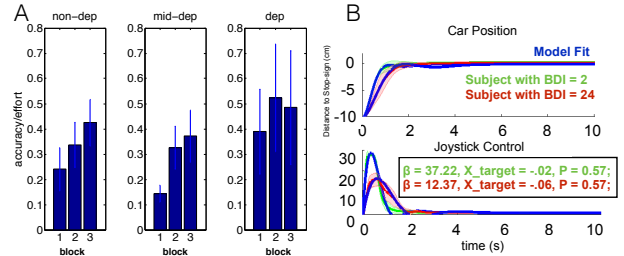


Figure 8: A: Model prediction of accuracy/effort ratio. B: Examples of model prediction. Those two subjects (healthy control $BDI = 0$ vs. depressed subject $BDI = 24$) have the same accuracy/effort ratio ($p = 0.57$), but the healthy subject has faster motor speed β and closer target distance X_{target} .

Discussion

In this paper, we proposed to use a simulated driving task and the inverse optimal control approach to examine the influence of depressed mood in motor-control in continuous time. We found depressed group has 1) slower perceptual and motor reaction time, 2) different behavioral goals but no significant difference in accuracy/effort ratio.

Our approach provided a computational framework to disentangle the factors between perceptual-motor speed and reward function in goal-directed motor-control tasks. The findings of slower perceptual-motor processing are consis-

tent with symptoms of psychomotor disturbance in depressed individuals (Treadway et al. 2009). Taking account of perceptual-motor speed in the feedback control loop, the findings of different reward-processing using inverse LQG model provided quantitative explanations of how different target states and target accuracy/effort ratio will influence motor-control in continuous time. However, these findings need to be interpreted with caution and require further investigation.

Target stopping distance If interpreting the intention of stopping further away from stop sign as to avoid crossing the stop-line, then our finding supports previous research showing depressed individuals have greater sensitivity to risk and punishment. However, it is important to consider other possible interpretations (Eshel & Roiser, 2010). In particular, one can argue that depressed individuals may have decreasing interest to perform the task due to anhedonia (Der-Avakian et al. 2012). Further research will be done to investigate this issue (risk-averse vs. disengagement from task due to anhedonia).

Target accuracy/effort ratio Our group-level comparison result suggests depression influences what goals individuals want to achieve, but not accuracy/effort ratio. This finding could imply depressed individuals may not necessarily have less willingness than non-depressed individuals to achieve their goals. Rather, the differences are in the choice of goals in a task. However, within depressed group, we also observed higher variability in both the goals and accuracy/effort ratio, which indicates high individual differences. Considering the many subtypes of depression, future research will be focusing on examining those individual differences, by considering other psychological factors (anxiety, personality traits, etc.) and use the model to further examine the relationship between perceptual-motor speed, goals and accuracy/effort ratio.

In conclusion, the combined behavioral and modeling approaches provide a tool to examine if and how the severity of psychomotor disturbance interacts with motivation deficits in depressed individuals.

References

- Beck, Steer, & Brown. (1996). *Manual for the beck depression inventory-ii*. Boston: Harcourt Brace.
- Bellman, R. (1957). *Dynamic programming* (1st ed.). Princeton, NJ, USA: Princeton University Press.
- Bulmash, E. L., Moller, H. J., Kayumov, L., Shen, J., Wang, X., & Shapiro, C. M. (2006). Psychomotor disturbance in depression: Assessment using a driving simulator paradigm. *Journal of Affective Disorders*, 93, 213 - 218.
- Buyukdura, J. S., McClintock, S. M., & Croarkin, P. E. (2011). Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 35(2), 395 - 409.
- Caligiuri, M. P., & Ellwanger, J. (2000). Motor and cognitive aspects of motor retardation in depression. *Journal of Affective Disorders*, 57(1-3), 83 - 93.
- Chapman, D. P., & Perry, G. S. (2008, January). Depression as a major component of public health for older adults. *Preventing chronic disease*, 5(1).
- Der-Avakian, A., & Markou, A. (2012, January). *The neurobiology of anhedonia and other reward-related deficits* (Vol. 35) (No. 1). Elsevier Applied Science Publishing.
- Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry*, 68(2), 118 - 124.
- Evans, L. (2004). *Traffic safety*. Science Serving Society.
- Hilton, M. F., Staddon, Z., Sheridan, J., & Whiteford, H. A. (2009, May). The impact of mental health symptoms on heavy goods vehicle drivers' performance. *Accident Analysis & Prevention*, 41(3), 453-461.
- Janssen, W. H. (1979). *Routeplanning en geleiding: Een literatuurstudie. report izf 1979 c- c13*. Soesterberg (NL): Institute for perception.
- Ng, A. Y., & Russell, S. (2000). Algorithms for inverse reinforcement learning. In *in proc. 17th international conf. on machine learning* (pp. 663-670). Morgan Kaufmann.
- Sabbe, B., Hulstijn, W., Hoof, J. van, Tuynman-Qua, H., & Zitman, F. (1999, September). *Retardation in depression: assessment by means of simple motor tasks* (Vol. 55) (No. 1). Elsevier/North-Holland Biomedical Press.
- Selzer, M. L., Rogers, J. E., & Kern, S. (1968). Fatal accidents: The role of psychopathology, social stress, and acute disturbance. *American Journal of Psychiatry*, 124(8), 1028-1036.
- Shadmehr, R., & Krakauer, J. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185(3), 359-381.
- Taylor, D. H. (1964). *Drivers' galvanic skin response and the risk of accident*. (Vol. 7) (No. 4). United Kingdom: Taylor & Francis.
- Todorov, E., & Jordan, M. I. (2002, November). Optimal feedback control as a theory of motor coordination. *Nat Neurosci*, 5(11), 1226-1235.
- Treadway, M. T., Buckholtz, J. W., Schwartzman, N., Ashley, Lambert, W. E., & Zald, D. H. (2009, 08). Worth the effort? the effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS ONE*, 4(8), e6598.
- Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, 31(7), 1156 - 1168.
- Wilde, G. J. S. (2002). Does risk homeostasis theory have implications for road safety. *BMJ*, 324(7346), 1149-1152.
- Wingen, M., Ramaekers, J., & Schmitt, J. (2006). Driving impairment in depressed patients receiving long-term antidepressant treatment. *Psychopharmacology*, 188(1), 84-91.