

# Goal-Subgoal Conflict as the Core Feature of Delay of Gratification

Violet Cheung-Blunden, Ph.D., and Jennifer L. Zahm, B.S.

## Abstract

*While delay of gratification is a well-established individual difference, its assessment is limited by the original delay paradigm, which is only valid in the age range of four to 11 years. If delay of gratification were to be construed as a competence of resolving goal-subgoal conflicts, a specific type of puzzle games, with a feature called ambiguous subgoal ordering, can be used as an alternative assessment tool. While the child version of this type of puzzle, the Dog-Cat-Mouse game (DCM), is available, an adult version, the Dog-Cat-Mouse-Rabbit game (DCMR), had to be developed. Two studies were conducted to establish the validity of DCM and DCMR (or DCM(R) in brief) by linking the performance in these games to delay time. To this end, 42 longitudinal participants from the original study of delay of gratification and 50 community children completed their age-appropriate versions of the planning task. Results showed that delay time was prospectively and concurrently associated with DCM(R) performance. Moreover, the mediator in the relationship was executive control (participants' success of negotiating goal-subgoal conflicts in the DCM(R) game) rather than intellectual ability (forward and backward digit span, the Peabody Picture Vocabulary Test or PPVT) in both the adult and child sample. These results offered support for DCM and DCMR as assessment tools for delay of gratification in children and adults, respectively. The DCMR's state space and its development are provided in the appendix.*

## Introduction

Delay of gratification refers to the ability to exercise self-control for the sake of a long-term goal (Mischel, 1974; Mischel, Gruesec & Masters, 1969; Mischel & Metzner, 1962; Mischel & Moore, 1973; Mischel, Zeiss, & Ebbesen, 1972). To assess this type of self-control, Mischel (1974) exposed four-year-old children to a dilemma where they had to choose between a small but immediate reward or a more substantial long-term reward. The amount of time that a preschooler was able to wait indexed his/her ability to delay gratification. Delay of gratification is one of many types of executive control (Aslan & Cheung-Blunden, 2012; Olson, Schilling, & Bates, 1999; White et al., 1994). The similarities among the subtypes have been the central tenet of Baumeister and colleagues (Gailliot & Baumeister, 2007), especially in terms of the shared glucose consumption and ego-depletion mechanisms.

A multimodal conception takes a different approach to the inter-relations among the subtypes of executive control. It emphasizes the distinctiveness of each subtype by pointing to

their various neurological substrates (Goel & Grafman, 1995). For example, the basal ganglia, a subcortical structure in the limbic system, has been associated with motor control, and has been shown to play a key role in disordered motor functioning, notably in Parkinson's disease, as well as in focal dystonia and stuttering (Obermann et al., 2008; Obeso et al., 2000). Alternatively, the prefrontal cortex, a cortical structure, is activated when individuals plan for the future and solve complex problems (Rowe, Owen, Johnsrude, & Passingham, 2001). The left prefrontal cortex is implicated in the negotiation of goal and subgoal conflicts whereas both left and right prefrontal cortices seem necessary in the Wisconsin Card Sort Task (Alvarez & Emory, 2006). Therefore, executive control is not a singular construct, but rather consists of several subtypes, including but not limited to, inhibition, mental set shifting, working memory monitoring (Miyake et al., 2000), motor control (Murray & Kochanska, 2002), attention sustention and shifting (Olson et al., 1999), cognitive inhibitory control (Kindlon, Mezzacappa, & Earls, 1995) and planning (Aslan & Cheung-Blunden, 2012).

Besides neurological substrates, other supporting evidence for the multimodal conceptualization of executive control is the developmental timeline (McCabe, Cunington, & Brooks-Gunn, 2004). While motor control and attention control emerge relatively early on, delay and dimensional sorting tend to appear around preschool (Rothbart, Sheese, Rueda, & Posner, 2011; Zelazo, Muller, Frye, & Marcovitch, 2003). The idea that all forms of executive control are integrally intertwined meets further resistance when considering some of the unique capabilities of adults. Self-monitoring (Snyder & Gangestad, 1986), for example, is the regulation of self-expression due to the demands and norms of an audience or context. Since a person's awareness of the self and the surrounding context is central to this form of executive control, the construct takes on a distinct nature and its onset is much later than other subtypes of executive control.

This multimodal conception of executive control is not evident in the longitudinal studies of delay of gratification thus far. There is little evidence suggesting an encapsulated form of executive control with a specific life outcome. Rather, delay has been linked to a wide array of executive control abilities in addition to advantages in other domains – delay score in preschool predicted multiple forms of executive control, emotion regulation, as well as intellectual competencies.

Specifically, several follow-up studies of the original sample found advantages with what might be considered common forms of executive control (inhibition, planning, attention, etc.). For example, a neurological study by Casey et al. (2011) followed up with Mischel's (1974) original preschoolers at 40 years of age and found activation differences in brain areas involved with inhibition. Additionally, increased drug use among those who scored low on the delay task (i.e. low delayers) (Wulfert, Block, Ana, Rodriguez, & Colman, 2002) and the opposite trend among those who scored high (i.e. high delayers) (Ayduk, Mendoza-Denton, & Mischel, 2000), offered further evidence for inhibition advantages associated with the ability to delay gratification in preschool. Developmental advantages in attention control and planning ability have also been found in the longitudinal follow-ups of the original participants. Adolescents who were high delayers in preschool were better able to plan ahead (Mischel, Shoda & Peake, 1988), and were described by their parents as having better attention control capabilities (Mischel, Shoda, & Rodriguez, 1989).

Besides common forms of executive control, another developmental advantage associated with longer delay time is emotion regulation, which is arguably a composite of more than one executive control ability (Ochsner & Gross, 2005). In a childcare setting, high delay scores

predicted more positive interactions during play with peers and higher maternal ratings of positive sociability (Ramani, Brownell, & Campbell, 2010). Verbal ability was found to buffer against aggressive behavior in elementary school children, but the buffering effect was stronger amongst those who were also high delayers (Ayduk, Rodriguez, Mischel, Shoda, & Wright, 2007). Additionally, parents of high delayers described their children as socially competent and equipped with adequate coping skills during adolescence (Mischel et al., 1989). In adulthood, high delay preschoolers have a reduced vulnerability to developing borderline personality features, which is a psychological disorder marked by severe emotional dysregulation (Ayduk et al., 2008).

Delay's predictive power reached the intellectual domain as well. For example, high delay preschoolers from Mischel's original preschool sample became adolescents who were more successful academically (Mischel et al., 1988), while low delay preschoolers demonstrated less academic achievement (Wulfert et al., 2002). They also obtained substantially higher SAT scores than their low delay peers in high school (Mischel et al., 1989) and eventually attained higher levels of education in young adulthood (Ayduk et al., 2000). Moreover, an international study conducted with ninth-grade Ireland students demonstrated that they received higher academic scores if they delayed longer in a task that was similar to the original delay paradigm (Freeney & O'Connell, 2010).

While a wide array of outcomes has ostensibly established delay of gratification as a classic construct in personality studies, the mechanism underlying the impressive predictive power is not well understood. A child could very well outshine his/her peers in a host of developments, and it is this general advantage rather than excellence in delay per se that continues to benefit him/her later in life (Wright, 2008). This kind of domain general explanation would be readily acceptable had there been a solid consensus behind a unidimensional structure of executive control. In other words, this line of longitudinal research assumes a certain degree of interchangeability/equivalence across the various subtypes of executive control.

As the concept of executive control expands to include emotion regulation and motivation on one end of the spectrum, and intellectual discipline and working memory on the other, some demarcations in the vast construct are theoretically meaningful. This notion is embedded in the multimodal conception of executive control, which challenges researchers to recognize each subtype of executive control for its distinctiveness. Before exploring the general value of possessing delay ability in related life outcomes, it is imperative to affirm the specific value of delay within its own domain. A domain specific research question can be posed around the core ability to delay gratification – whether participants, who demonstrated competency in delaying gratification at some point, can once again forego the attractive short-term goal and obtain the more substantial long-term goal when given another chance to resolve goal tradeoffs. This research question can be pursued by cross-sectional studies where participants delay gratification in multiple settings, or by longitudinal studies where participants exhibit their delay ability across time.

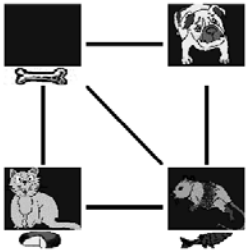
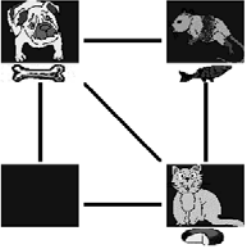
To answer this research question, a test-retest design is necessary. However, a roadblock has been the limited validity of the original delay paradigm, which is only applicable in the age range of four to 11 years (Rodriguez, Mischel, & Shoda, 1989). Although some research effort was devoted to devising another delay-like situation for adults (Kirby & Marakovic, 1996), the face values of different rewards in children and adults cast doubt on the construct validity.

The present study leverages a special class of puzzle games and argues that their ambiguous goal-subgoal ordering affords participants similar opportunities to negotiate the

tradeoff between two temporal goals. We utilize a preexisting child version of the game and a newly developed version for adults. We examine the concurrent and prospective relationships between delay time and the performance on these puzzle games while teasing out the effect of intellectual capacities to show that it is the core ability of delay that is responsible for our findings in the cross-sectional and longitudinal studies. We argue that these puzzle games can overcome the age barrier of the classic delay paradigm and recommend their use as an alternative assessment tool for the core delay ability, which is defined as the ability to forego an attractive short-term goal in order to obtain a more substantial long-term goal.

The Dog-Cat-Mouse (DCM) game was developed by Borys (1984). Klahr (1985) suggested that the DCM games constitute a special class of puzzle games due to their ambiguous subgoal ordering. Ambiguous subgoal ordering means that there are several potential subgoals a player can implement and the correct choice is not immediately apparent. In fact, a handful of DCM games tempt players to make erroneous moves that change the current configuration to superficially resemble the goal configuration (Anderson, 1982; Klahr, 1985). However, in reaching the wrong subgoal (an immediate animal/food match), the player diverts from the ultimate goal (matching all animals to their foods). This type of impulsive move is called hill-climbing and it tends to undercut overall game performance. For example, impulsively moving the dog to the bone in the first DCM game in Table 1, when the player should move the cat first, results in a higher hill-climbing score; so does keeping the dog with the bone when the dog should be moved out of the way in the second DCM game in Table 1. These short-sighted moves will prevent the player from solving the puzzle games efficiently as specified by the optimum solution length (i.e. minimum number of moves to reach the ultimate goal).

Table 1  
DCM and DCMR Games in the Present Study

		Starting State	Optimum	Critical
		Ending State	Solution	Juncture
Child DCM				
1 <sup>st</sup> game		2 18	4	1
2 <sup>nd</sup> game		11 20	5	1

3 <sup>rd</sup> game		13 19	6	1
<hr/>				
Adult DCMR				
1 <sup>st</sup> game		X-5 III-2	6	1
2 <sup>nd</sup> game		III-1 II-2	7	2
3 <sup>rd</sup> game		VII-5 VIII-5	9	1

*Note.* The starting and ending states are numbered according to the state diagrams (see Figure 3 in Klahr, 1985, for DCM and Figure 1 in the present study for DCMR). Critical juncture indicates the step where hill-climbing opportunities are offered.

The technical aspects of developing a more advanced version of DCM, the Dog-Cat-Mouse-Rabbit (DCMR) games can be found in the Appendix. Briefly, all DCMR games in Table 1 offer hill-climbing opportunities, and when a person succumbs to these temptations – refuses to move the cat away from the fish in the first DCMR game in Table 1 or moves the rabbit prematurely to the carrot in the second DCMR game in Table 1 – the optimum solutions are out of reach.

As planning tasks, DCM(R) games also require a person to retain a certain number of potential actions, traverse across competing solutions, select the optimum solution and go backward to the first step to implement that particular solution (Klahr, 1985). Indices of IQ, such as digit span and PPVT have been established as component skills of planning both theoretically

and empirically (Carlson, Moses, & Claxton, 2004; Charness, 1976; Hudson, Shapiro, & Sosa, 1995; Spitz & Borys, 1984). Even though the performance of DCM(R) may be a function of IQ, delay has to also be predicated on IQ in order for the intellectual component to be a key mechanism in the relationship between delay and DCM(R).

### The Present Study

The present study aims to link the ability to delay gratification to the performance on DCM(R) and seeks to replicate the result in two age groups. The first group consisted of a convenience sample of community children who were in the valid age range to complete the delay task and DCM game. The second group was comprised of adults. While the DCMR game was age-appropriate for them, the original delay task was not. Hence, we recruited adults from a longitudinal database in order to retrieve their delay scores from record. We hypothesized a prospective and a concurrent relationship between delay time and scores on the DCM(R) game. We further hypothesized that the core competence in delay – ability to forego an attractive short-term goal in order to obtain a more substantial long-term goal – would emerge as a significant mediator in the relationship. Although IQ has been implicated in past studies of delay, it is, theoretically, an auxiliary component of delay and thus will not be a significant mediator in the concurrent or in the prospective relationship.

### Study 1

A group of longitudinal participants from the original study of delay of gratification was recruited in order to investigate the prospective relationship, as well as the underlying mechanism, between preschool delay and DCMR performance at 40 years of age.

### Method

**Participants.** The adult sample included 42 participants ( $M=40.30$  years,  $SD=1.97$ ) from the original delay of gratification study. The sample was 38.1% male and 61.9% female. All participants held a college degree, with 31.7% further completing a master's degree and 12.2% further completing a doctorate degree. The mean standardized score ( $M=112.80$ ,  $SD=11.03$ ) on PPVT was significantly higher than the population mean of 100 ( $t(42)=65.49$ ,  $p < .000$ ). The sample was ethnically homogenous with principally Caucasians born in the U.S.

**Procedure.** The adult participants were selected from the longitudinal database managed by Walter Mischel and colleagues (Mischel et al., 1989). Despite the one-third attrition rate after 35 years, the present study was able to recruit from the remaining participants (67.8%) with two selection criteria – married with at least one child between the age of four and 11. These selection criteria were not relevant to this study, but were adopted for the purpose of investigating the intergenerational transmission of delay/self-control in a different study. After a letter explaining the purpose of the research and a phone call to schedule a visit, the rate of participation from those who were contacted was 67.2%.

The adults' delay scores were measured in preschool (see Shoda, Mischel & Peake, 1990; Ayduk et al., 2000 for detailed descriptions). The present study followed the same procedure used by previous studies, where the delay time of each participant was subtracted by the mean delay time of the relevant condition. The delay time of adults who participated ( $M=0.64$  minutes,  $SD=5.39$  minutes) and those who did not participate ( $M=-0.71$  minutes,  $SD=6.03$  minutes) were

not significantly different ( $t(63)=1.40, p=0.166$ ).

Human subject approval was obtained for all experimental procedures including administering IQ tests and computer games (for the present study). Participants signed the consent forms and completed a demographic questionnaire, sub-scales of IQ tests, and the DCMR game.

**Instruments.** A demographic questionnaire was used to gather background information such as birthday, sex, ethnicity, birth place, education level, and marital status.

The Peabody Picture Vocabulary Test 3<sup>rd</sup> ed. (PPVT-III) was used to assess receptive vocabulary and the Wechsler Adult Intelligence Scale, 3<sup>rd</sup> ed. (WAIS-III) was used to measure forward and backward digit spans.

Three DCMR games (Table 1) were administered to adults. The games were implemented as web-based applications (i.e., an applet) using the Java runtime environment from Sun Microsystems (Blunden, 2005). Instructions on the computer were similar to Klahr (1985) and Spitz and Borys (1984), which informed participants that they would play each game either once or twice depending on whether the optimum solution was found on the first trial. For each game, the applet tracked the user's journey and recorded the pathways from start to finish. Participants were informed of their total number of steps at the end of each trial in a dialog box. If the length of the path was greater than the optimal value, the user was prompted to play the same game over again if it was the first trial or move on if it was the second trial. After all games were completed, the applet sent the user's results to a file, where the data was stored and coded.

The computer program coded the planning performance on a game-by-game basis: A score of "2," "1," or "0" was assigned to each game depending on whether the optimum solution was discovered on the first try, the second try, or neither (Klahr, 1985; Spitz & Borys, 1984). An overall performance score was then computed by averaging the scores across the games. A binary hill-climb score was coded at the critical junctures (where an opportunity for hill-climbing was present) for each trial. This coding system ensured that "1" stands for hill-climbing (prematurely matched an animal to its food, or a reluctance to give up an animal-food match) and "0" stands for the inverse (Klahr, 1985). An average hill-climbing score was computed by aggregating across the total number of trials.

**Models and Statistical Techniques.** The steps of mediation analysis recommended by Baron and Kenny (1986) were followed. Our independent variable was delay time and the dependent variable was DCMR performance score. There were two sets of competing mediators, a) intellectual capacities such as PPVT score, forward and backward digit span, and, b) hill-climbing score. For example, in order to examine hill-climbing as a mediator, three regression models were examined:

$$\text{Model 1: DCMR Performance} = \gamma_1 + \tau \text{Delay} + \varepsilon_1$$

$$\text{Model 2: Hill-climbing} = \gamma_2 + \alpha \text{Delay} + \varepsilon_2$$

$$\text{Model 3: DCMR Performance} = \gamma_3 + \tau' \text{Delay} + \beta \text{Hill-climbing} + \varepsilon_3$$

Instead of using linear regression, we employed hierarchical linear models (HLMs) to implement the mediation analysis. This method offers a way to keep all data points by recognizing that the three DCMR games played by the same participant are related to each other. This application of HLM is not only in line with its general function, which is to manage observations that fall into groups or clusters (Rabe-Hesketh & Skrondal, 2005; Rabe-Hesketh, Skrondal, & Pickles, 2002), but also brings the critical advantage of enhanced statistical power to

compensate for the small sample size. Rather than 42 performance scores, the HLM method analyzed 126 performance scores as each participant actually played three DCMR games and had three scores. Since the analysis has arrived at the game level, we also controlled for game characteristics (such as solution length and game sequence) in order to obtain more concise estimates for the key variables (independent and mediators) of interest.

**Results**

The inter-correlations among the main variables in the adult sample can be found in Table 2. Our hypothesis on the prospective relationship was supported. Delay score in preschool was predictive of DCMR performance at the age of 40. Longer delay time in preschool was associated with higher DCMR scores 35 years later.

Table 2  
*Intercorrelations among Main Variables in the Adult Sample*

	1	2	3	4	5	6
1. Age						
2. Forward Digit	-.29					
3. Backward Digit	-.29	.58***				
4. Standardized PPVT	-.30	.18	.25			
5. Preschool Delay <sup>a</sup>	.16	-.11	-.17	.26		
6. DCMR Performance	.14	.35*	.29	.20	.32*	
Mean	40.32	11.53	8.10	112.83	0.64 <sup>a</sup>	1.31
SD	1.99	2.28	2.56	12.03	5.39	0.51
Minimum	36.83	7	4	88	-8.58	0
Maximum	44.17	16	14	139	8.67	2
Scale	--	0 – 16	0 – 14	--	--	0 – 2

Note. <sup>a</sup> Preschool delay time is the deviation from the mean in minutes. \* $p < .05$ , \*\*\* $p < .001$ .

We explored the possibility of intellectual components (i.e. PPVT and digit span scores) serving as mediators in the longitudinal relationship but the premise required by Baron and Kenny (1986) was missing. Especially noteworthy in Table 2 was the lack of significant correlations between the independent variable (delay time) and the mediators (PPVT and digit spans). The correlations between the mediators (PPVT and digit spans) and the dependent variable (DCMR performance) were not consistently significant either. Therefore, intellectual competency was unable to explain the high delay preschoolers’ ability to effectively solve DCMR games in middle adulthood.

We proceeded to examine hill-climbing as a mediator in the prospective relationship between preschool delay and adult DCMR performance. HLMs were used and the results can be found in Table 3. Following the steps recommended by Baron and Kenny (1986), Model 1 examined the relationship between the independent variable and the mediator. Longer delay time in preschool was found to have a significant positive effect on reducing hill-climbing moves 35 years later ( $B=-0.01, p=.047$ ). Solution length and game numbers were included as predictors to control for game difficulty and practice effects. The rest of the rows in Table 3 are necessary parameters in HLM, akin to the idea of intercepts in regular regression models.



Table 3  
*Four Models Examining the Mediating Effect of Hill-climbing in the Relationship between  
 Preschool Delay Time and Adult DCMR Performance*

	Model 1 for Hill-climb <sup>a</sup>	Model 2 for DCMR <sup>a</sup>	Model 3 for DCMR <sup>a</sup>	Model 4 for DCMR <sup>a</sup>
	<i>B</i> ( <i>SE</i> )	<i>OR</i> (95% CI)	<i>OR</i> (95% CI)	<i>OR</i> (95% CI)
Delay	-0.01* (0.00)	--	1.07* (1.00, 1.15)	1.06 (0.98, 1.14)
Hill-climb	--	0.00*** (0.00, 0.00)	--	0.00*** (0.00, 0.00)
SL <sup>a</sup>	0.05* (0.02)	0.82 (0.51, 1.32)	0.83 (0.62, 1.11)	1.20 (0.80, 1.83)
Game 2	0.07 (0.07)	1.22* (1.02, 1.46)	0.63 (0.29, 1.37)	0.87 (0.32, 2.39)
Threshold 1	--	-3.05	-2.71	-3.22
Threshold 2	--	-0.37	-1.72	-0.36
$\sqrt{\varphi}$	0.05	--	--	--
$\sqrt{\theta}$	0.33	--	--	--

Note. SL – solution length. <sup>a</sup> The dependent variable for Model 1 is hill-climbing, and for Models 2, 3 and 4 is DCMR performance. Game 3 dropped due to co linearity. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

The remaining models (columns) in Table 3 are proportional odds models, such that the odds ratio of one refers to equal odds. For example, Model 2 examined the relationship between the mediator (hill-climbing score) and the dependent variable (DCMR score). The odds of solving a DCMR game successfully were small, should one decide to hill-climb. Model 3 examined the relationship between the independent (delay time) and dependent (DCMR score) variables. Longer delay time in preschool placed the odds of solving the DCMR game 35 years later significantly above one. Therefore, had a preschooler waited one extra minute in the classic delay paradigm, his/her odds of solving the DCMR game successfully would be 1.07 times ( $p=0.048$ ) his/her original odds.

Model 4 pivoted the independent variable (delay time) against the mediator (hill-climbing score) to jointly explain the dependent variable (DCMR score). Results showed that delay was no longer a significant predictor ( $p=0.160$ ) whereas hill-climbing remained a significant ( $p<.000$ ) predictor for DCMR performance. It appears as though preschoolers who were better at delay of gratification developed into adults with greater skills for negotiating the goal-subgoal conflict (i.e. had a lower hill-climbing score), which, in turn, enhanced their DCMR performance.

## **Study 2**

A group of community children was recruited in order to investigate the concurrent relationship, as well as the underlying mechanism, between delay time and DCM performance.

### **Method**

**Participants.** The child sample included 50 children who ranged in age from four to eight years old ( $M=6.99$  years,  $SD=1.18$  years). There were more boys (70.0%) than girls (30.0%). Their ethnic backgrounds were primarily Caucasian (62.0%), followed by 8.0% Latino, 6.0% Asian, 4.0% African-American, 4.0% Middle-Eastern and 16.0% other. A great majority was the oldest child in the family (83.0%) and the rest (17.0%) were the second child. More than one-third (36.7%) of the children had no siblings, whereas 51.0% had one and 12.2% had two siblings.

**Procedure.** Children were recruited via advertising and referrals. Human subject approval was secured for all procedures including administering food to children, video recording their behaviors during sessions of delay and DCM games, and assessing receptive vocabulary and digit spans. After obtaining consent and demographic information from the parents, researchers administered the delay task, sub-scales of IQ tests, and the DCM game to child participants.

**Instruments.** The demographic questionnaire and the Peabody Picture Vocabulary Test were the same as Study 1. The Wechsler Intelligence Scale for Children 4<sup>th</sup> ed. (WISC-IV) was used to measure forward and backward digit spans.

In the delay task, children were given a choice between two types of food and then read the delay instructions consistent with the original study (Shoda et al., 1990). Briefly, a child was seated in a chair and a counter bell was placed on the side of a child's dominant hand on a table in front of him/her. The child was shown how the bell rang and its function in bringing the researcher back to the room. The child tried the bell twice and saw the researcher return. The child was presented with a choice between pretzels and cookies, and then a choice between one or two of their preferred treat. A paper plate containing three of their preferred treats was placed on the table, with one treat on the dominant hand side, and two on the non-dominant side. The researcher then explained the rules – two treats could be obtained by waiting the whole time for the researcher to come back in their own accord and one treat could be obtained by ringing the bell to bring the researcher back. The child was asked questions to ascertain their willingness to wait and their understanding of the rules. The child was also told to not leave the seat or touch the food. The delay task was video recorded and the delay time was stopped if they touched the food, left the seat, rang the bell or if the researcher returned. Delay time was measured in seconds and inter-rater reliability was .95.

The child participants received a board game version of the DCM game, which was constructed by a woodshop. The way to contextualize this game for children is to use three animals to specify the initial state, their favorite foods to specify the goal state, and interconnected grooves to specify valid paths (Table 1). Instructions, similar to those in Klahr (1985), were read aloud by a researcher. The game was described using a story, which was about getting a dog, a cat, and a mouse to their favorite foods (i.e. a bone, a fish, and a piece of cheese). The movement of the animals was subject to the constraints that: a) only one animal could be

moved at a time, and, b) only one animal could occupy any given vertex at a time (Borys, 1984; Klahr, 1985). Verification questions were asked in order to ascertain children’s understanding of the rules: “Which food does the dog want?” “Which food does the cat want?” “Which food does the mouse want?” “Can two animals move at the same time?” “Can two animals sit in the same corner at the same time?” Children were given further explanations until they could answer all questions correctly. A practice game was administered, followed by the three test games in Table 1. If the length of the path was greater than the optimal value, participants were informed of their total number of steps and were prompted to play the same game again. At the end of the second try, participants moved onto the next game regardless of performance. The games were videotaped for later coding by two coders. The overall performance and hill-climbing scores were coded using a similar scheme as Study 1. The performance on each game was scored as “2,” “1,” or “0,” depending on whether the optimum solution was reached on the first try, second try, or neither; hill-climbing was scored as “0” for not hill-climbing and “1” for hill-climbing at the critical juncture for each try. A total performance score and an average hill-climbing score were calculated accordingly. Since the coding involved simple counting, the two coders were in complete agreement.

**Models and Statistical Techniques.** The mediation hypotheses were examined by following the steps recommended by Baron and Kenny (1986). Bivariate correlations were used to implement the first three steps, with the aim to ascertain significant associations among the independent variable (delay score), mediators (PPVT, digit spans, and hill-climbing) and dependent variable (DCM score). Linear regression was used to implement the last step, with a focus on the predictive power of the mediators (PPVT, digit spans, and hill-climbing) over the predictive power of the independent variable (delay time) when both were entered into a regression analysis for the dependent variable (DCM score).

Table 4  
*Intercorrelations among Main Variables in the Child Sample*

	1	2	3	4	5	6	7
1. Age							
2. Forward Digit	.36**						
3. Backward Digit	.60***	.45***					
4. Standardized PPVT	-.08	.07	.28				
5. Preschool Delay	.35**	.34*	.38**	.15			
6. Hill-climb	-.17	-.14	.30*	.08	-.31*		
7. DCM Performance	.27	.20	.25	.07	.34*	-.65***	
Mean	6.99	8.19	3.73	115.18	339.98	0.61	0.80
SD	1.18	2.11	1.61	15.43	838.92	0.32	0.48
Minimum	4.33	4	0	80	0	0	0
Maximum	8.83	14	8	144	1804.8	2	2
Scale	--	0 – 16	0 – 14	--	--	0 – 2	0 – 2

Note. <sup>a</sup> Preschool delay time is the deviation from the mean in minutes. \* $p < .05$ , \*\*  $p < .01$ , \*\*\* $p < .001$ .

## Results

Before conducting the mediation analyses, the effect of demographics on DCM game performance was examined. Results showed that performance on the DCM game did not seem to differ by age ( $r=.27$ ,  $p=.068$ ), number of siblings ( $r=-.02$ ,  $p=.883$ ), birth order ( $r=.13$ ,  $p=.414$ ), ethnicity ( $F(5, 47)=1.90$ ,  $p=.114$ ) or sex ( $t(46)=1.08$ ,  $p=.285$ ).

Without having to control for demographics, the first three steps of the mediation analyses were carried out by bivariate correlations (Table 4). Results showed that the independent variable (delay score) and dependent variable (DCM score) were significantly correlated. However, the IQ measures (PPVT and digit spans) were not significantly correlated with the dependent variable (DCM score). As such, our results did not support IQ as a mediator in the relationship between delay of gratification and DCM game performance.

Hill-climbing showed significant correlations with both delay and DCM score, and thus was further analyzed as a mediator. With delay as the predictor for DCM score, the regression coefficient was  $.34$  ( $p=.020$ ). When hill-climbing was also entered into the regression model, the coefficient of delay was no longer significant ( $\beta=.16$ ,  $p=.192$ ) and the coefficient of hill-climbing was significant ( $\beta=-.61$ ,  $p=.000$ ). This finding lent support to our mediation hypothesis that the reason high delay children were able to score higher on the DCM game was because they were better able to resist hill-climbing.

## Discussion

By following up with the original delay sample and examining a group of community children, the present study found a prospective as well as a concurrent correlation between delay time and performance on the DCM(R) game. Additionally, results provide evidence for a specific executive control advantage, which fits with the core definition of delay of gratification. Our mediation analysis pointed to hill-climbing, rather than intellectual abilities (short-term memory and perceptive vocabulary) as the main reason why high delayers excelled in planning amongst both adult and child samples. Specifically, an inability to wait during the delay task predicted an increased number of hill-climbing moves on the DCM(R) game. For example, if an adult participant had been able to wait one additional minute in the delay paradigm during preschool, his/her odds of successfully solving the DCMR game in adulthood would increase about 1.07 times. The mediator is that the person is less inclined to use hill-climbing strategies (by approximately one less strategy per 100 DCMR games). Moreover, our investigation of the mechanism underlying the relationship between delay time and DCM performance in the child sample corroborated our conclusion from the longitudinal sample.

Previous studies found high delayers to enjoy a wide array of developmental advantages in terms of intellectual competency and multiple executive controls, but the present study focused on the core concept of delay of gratification. The mediator, hill-climbing, provided a concise view of this core concept – the ability to forego a small immediate goal for a large long-term goal was consistent in two situations that were administered simultaneously for children and 35 years apart for adults. These findings serve to show that the core ability to delay gratification is expressed across task settings and over a time span of 35 years.

The tool that allowed us to visit the research question about the core ability of delay is a class of planning tasks with ambiguous subgoal ordering. When it is not apparent to a player which subgoal to implement, he/she can easily fall for the attractive immediate subgoal. A careful selection of the games ensures that the gravitation towards the attractive subgoal is wrong and will cost the player the overall goal. This kind of goal-subgoal conflict can be framed in the

language of delay – the overall goal is large and distant whereas the subgoal is small and immediate. In other words, a small immediate goal (one animal gets to eat) is pivoted against a large long-term goal (all animals get to eat). This feature resembles the situation of the classic delay paradigm where the choice is between one treat now and two treats later. Therefore, the goal-subgoal conflict invites the expression of a particular type of impulsiveness in a similar way as the classic delay paradigm.

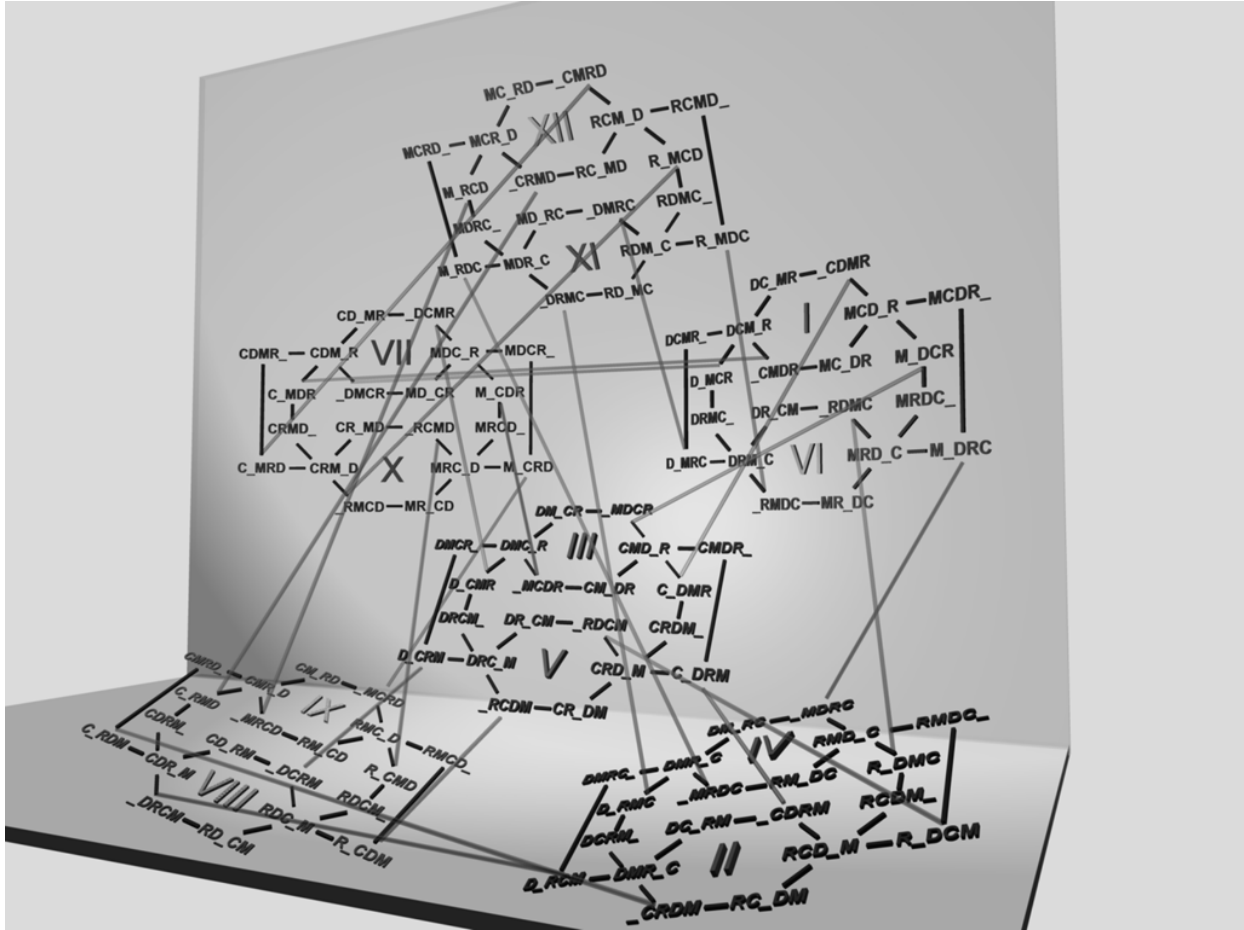
Klahr (1985) used a handful of DCM games that present goal-subgoal conflicts to children. Following the established protocol for game development (Appendix), we utilized a more advanced class of DCM-type problems for adults, which adds a game piece and hence increases the degrees of freedom of movement. The new four-object, five-vertex version, called DCMR, has an expanded state space. The game's state diagram (Figure 1) is the master key to all DCMR games and a design tool to locate a DCMR game with desired characteristics. DCMR games have all the inherent characteristics of DCM including the goal-subgoal conflicts or the opportunities for participants to hill-climb.

Since hill-climbing is the empirical and theoretical link between delay and DCM(R), we suggest using DCM and DCMR to approximate the ability to delay gratification in studies where a) food or video-recording are impractical, or, b) the age of participants are outside the valid range of the classic delay paradigm. The key metric could be DCM(R) performance score or hill-climbing score.

Although benchmarking is not a goal of our study, we can glean for age-appropriateness from the performance scores. For example, the mean performance of our child sample on the DCM games was 1.31 with a standard deviation of 0.51 (Table 1) on a scale of 0 to 2. Since the mean is a little less than two standard deviations from the top score of 2, a slight ceiling effect may exist. Future studies working with older school aged children may consider adding a DCM game with an optimum solution length of seven or eight steps, where eight is the longest game allowed by DCM state space (Klahr, 1985). Our adult sample achieved a mean score of 0.8 with a standard deviation of 0.32 on DCMR games. Since the mean fits within plus or minus two standard deviations within the range of 0 to 2, no floor or ceiling effect was evident. The DCMR games in Table 1 may be used in future studies of adults.

The present study did not include adolescents because a sample of adolescents with delay scores on record was unavailable to us. Future studies of adolescents will have to recruit a cohort of four to 11-year olds, administer the classic delay paradigm, wait until they reach adolescence, and then administer DCMR games. Short of such longitudinal efforts, we can rely on the principle of linear extrapolation to recommend age-appropriate games for adolescents. Linear extrapolation is a feasible principle due to the precedence set by Tower of Hanoi (see appendix) which means that optimal solution length marks small increments in difficulty level and game version marks a larger leap in difficulty level. In fact, if certain applications call for a direct comparison of DCM and DCMR scores, optimum solution length and game versions are the parameters to control for. This kind of quantification is not possible without the well-defined formal property and the state diagrams of the DCM(R) games. Based on the performance of the adult and child samples in the present study, we recommend to use the easier games of the hard version for adolescents, i.e. DCMR games with a solution length of 4, 5, or 6 (see appendix for procedures of finding DCMR games from its state diagram; the first DCMR game in Table 1 may also be appropriate with its solution length of 6).

Figure 1  
The State Diagram of DCMR



Note. Each state is denoted by a five-letter string with the positions in the string corresponding to the actual positions on the game board: the first letter represents the upper left vertex; the second, upper right; the third, lower left; the fourth, lower middle; and the last letter represents the lower right vertex. For example, the starting state of the first DCMR game in Table 1 is denoted by MR\_CD and the end state is denoted by DM\_CR. An accepted practice is to name/number each state on a state diagram. Our nomenclature is to use Roman numerals to refer to regions of the state diagram and integer values to specify the position of a state in a region. Thus, each state can be denoted by the concatenation of a Roman numeral and an integer value. There are six states immediately surrounding each Roman numeral and they are numbered in a clockwise fashion starting from the middle left position. For example, DMC\_R, DM\_CR, \_MDCR, CMD\_R, CM\_DR, and \_MCDR are labeled III-1 to III-6 sequentially. There are eight states between each pair of Roman numerals and they are numbered from top to bottom, and then laterally. For example, DMCR\_, D\_CRM, D\_CMR, DRCM\_, C\_DMR, CRDM\_, CMDR\_, and C\_DRM are identified by III-V-1 to III-V-8 respectively.

## **Limitations**

Several caveats are noteworthy in the present study. First, we acknowledge that while we have attempted to single out a specific type of executive control that mirrors the core construct of delay, various executive controls are necessarily linked to one another and there are likely influences beyond the ability to negotiate short versus long-term goals. However, we believe that hill-climbing serves to refocus the current literature in that it seems to be theoretically close to the definition of delay.

Other limitations include that the longitudinal sample was not representative of the general population given the predominantly Caucasian ethnic composition, the use of a preschool at a university to recruit participants, and the high SES of this sample. To this end, we included a cross-sectional sample of children from the community to increase the generalizability of our findings. Future studies should strive for diverse samples to further the generalizability of these findings.

A related issue that is common in longitudinal studies is retention, where a long time span often results in small sample size. As a remedy, we increased the statistical power by asking each participant to contribute multiple data points and used HLM to manage the relatedness among the data points. Another precaution we took was to ascertain that the delay scores of those who participated and those who did not participate in the study were not significantly different. Even if there is selective dropout, the textbook argument does not only pertain to representativeness but also to the reduced variability within the sample. Previous studies have shown that executive control varies a great deal even among high delayers. When the heart rate reactivity (HRR) and electro dermal responding (EDR) of children completing the delay task were compared, physiological differences were observed (Wilson, Lengua, Tininenko, Taylor, & Trancik, 2009). Among children who were able to wait the entire time, the cluster with low HRR/low EDR, exhibited less inhibitory and attention control, and reported depression and conduct related problems years later. Therefore, it is important to remember that high delayers and likely, low delayers are not homogeneous groups and that factors beyond executive control abilities likely influence outcomes as well.

## **About the Authors**

**Violet Cheung-Blunden, Ph.D.** is an assistant professor in the psychology department at the University of San Francisco. Her research addresses emotions and emotion regulation in basic and applied contexts.

**Jennifer L. Zahm, B.S.** is a graduate student obtaining her doctoral degree in Clinical Psychology from Palo Alto University's Pacific Graduate School of Psychology. She is in her third year of the program and her current research endeavors involve examining anxiety and inflammation in the Trauma and Health Research on Immunity, Vitality and Emotions (THRIVE) lab at the University of California, San Francisco/San Francisco Veterans Affairs Medical Center.

## **Acknowledgement**

We wish to thank the University of San Francisco, Bill Blunden, Michael Vilensky, Ozlem Ayduk, Sophia Rabe-Hesketh, Lori Markson, Stephanie Carlson, as well as the support from researchers associated with the original study of delay of gratification.

## References

- Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*, 17-42.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.
- Aslan, S., & Cheung-Blunden, V. (2012). Where does self-control fit in the Five-factor Model? Examining personality structure in children and adults. *Personality and Individual Differences, 53*, 670-674.
- Ayduk, Ö., Mendoza-Denton, R., & Mischel, W. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. *Journal of Personality and Social Psychology, 79*, 776-792.
- Ayduk, Ö., Rodriguez, M. L., Mischel, W., Shoda, Y., & Wright, J. (2007). Verbal intelligence and self-regulatory competencies: Joint predictors of boys' aggression. *Journal of Research in Personality, 41*(2), 374-388. doi: 10.1016/j.jrp.2006.04.008
- Ayduk, Ö., Zayas, V., Downey, G., Cole, A. B., Shoda, Y., & Mischel, W. (2008). Rejection sensitivity and executive control: Joint predictors of borderline personality features. *Journal of Research in Personality, 42*(1), 151-168. doi: 10.1016/j.jrp.2007.04.002
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Blunden, B. A. (2005). *A breath-first search algorithm of the Dog-Cat-Mouse-Rabbit puzzle*. (Unpublished manuscript). University of California at Davis, Davis, CA.
- Borys, S. V. (1984). *Depth of search capacity: A developmental study*. (Unpublished doctoral dissertation). Graduate School of Education, Rutgers, the State University of New Jersey, NJ.
- Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Child Psychology, 33*, 87-110.
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299-319.
- Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., Jonides, J., Berman, M. G., Wilson, N. L., Teslovich, T., Glover, G., Zayas, V., Mischel, W., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later.



- PNAS Proceedings of the National Academy of Sciences of the United States of America*, 108(36), 14998-15003. doi: 10.1073/pnas.1108561108
- Charness, N. (1976). Memory for chess positions: Resistance to inference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641–653.
- Cheung, V. (2007). *Delay of gratification and strategic planning*. (Unpublished doctoral dissertation). University of California at Berkeley, Berkeley, CA.
- Freeneey, Y., & O'Connell, M. (2010). Wait for it: Delay-discounting and academic performance among an Irish adolescent sample. *Learning and Individual Differences*, 20(3), 231-236. doi: 10.1016/j.lindif.2009.12.009
- Gailliot, M. T., & Baumeister, R. F., (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review*, 11, 303-328.
- Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in 'planning' functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33(5), 623-642. doi:10.1016/0028-3932(95)90866-P
- Hudson, J. A., Shapiro, L. R., & Sosa, B. B. (1995). Planning in the real world: Preschool children's scripts and plans for familiar events. *Child Development*, 66, 984-998.
- Kindlon, D. J., Mezzacappa, E., & Earls, F. (1995). Psychometric properties of impulsivity measures: Temporal stability, validity and factor structure. *Journal of Child Psychology and Psychiatry*, 36, 645-661.
- Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin and Review*, 3, 100-104.
- Klahr, D. (1985). Solving problems with ambiguous subgoal ordering: Preschoolers' performance. *Child Development*, 56, 940-952.
- McCabe, L., Cunnington, M., & Brooks-Gunn, J. (2004). The development of self-regulation in young children: Individual characteristics and environmental contexts. In R. F. Baumeister, & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 340-356). New York: Guilford Press.
- Mischel, W. (1974). The layman's use of trait and behavioral information to predict behavior. *Journal of Research in Personality*, 8, 231-242.
- Mischel, W., Grusec, J., & Masters, J. C. (1969). Effects of expected delay time on the subjective value of rewards and punishments. *Journal of Personality and Social Psychology*, 11(4), 363-373. doi: 10.1037/h0027265

- Mischel, W., & Metzner, R. (1962). Preference for delayed reward as a function of age, intelligence and length of delay interval. *Journal of Abnormal and Social Psychology*, 64(6), 425-431. doi:10.1037/h0045046
- Mischel, W., & Moore, B. (1973). Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology*, 28, 172-179.
- Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, 54, 687-696.
- Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, 244, 933-938.
- Mischel, W., Zeiss, R., & Ebbesen, E.B. (1972). *Specificity and consistency in delay of gratification*. (Unpublished manuscript). Stanford University, Palo Alto, CA.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.
- Murray, K., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*, 30, 503-514.
- Obermann, M., Yaldizli, O., de Greiff, A., Konczak, J., Lachenmayer, M. L., Tumczak, F., Buhl, A. R., Putzki, N., Vollmer-Haase, J., Gizewski, E. R., Diener, H.C., & Maschke, M. (2008). Increased basal-ganglia activation performing a non-dystonia-related task in focal dystonia. *European Journal of Neurology*, 15, 831-838.
- Obeso, J. A., Rodriguez-Oroz, M. C., Rodriguez, M., Lanciego, J. L., Artieda, J., Gonzalo, N., & Olanow, C. W. (2000). Pathophysiology of the basal ganglia in Parkinson's disease. *Trends in Neurosciences*, 23, S8-S19.
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242-249. doi: <http://dx.doi.org/10.1016/j.tics.2005.03.010>
- Olson, S., Schilling, E., & Bates, J. (1999). Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology*, 27, 151-165.
- Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, Texas: StataCorp LP.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal*, 2, 1-21.

- Ramani, G. B., Brownell, C. A., & Campbell, S. B. (2010). Positive and negative peer interaction in 3- and 4-year-olds in relation to regulation and dysregulation. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *171*(3), 218-250. doi: 10.1080/00221320903300353.
- Rodriguez, M. L., Mischel, W., & Shoda, Y. (1989). Cognitive person variables in the delay of gratification of older children at risk. *Journal of Personality and Social Psychology*, *57*, 358-367.
- Rothbart, M. K., Sheese, B. E., Rueda, M., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion Review*, *3*(2), 207-213. doi:10.1177/1754073910387943
- Rowe, J. B., Owen, A. M., Johnsrude, I. S., & Passingham, R.E., (2001). Imaging the mental components of a planning task. *Neuropsychologia*, *39*, 315 – 327.
- Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*, 978-986.
- Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, *51*(1), 125-139. doi:10.1037/0022-3514.51.1.125
- Spitz, H. H., & Borys, S. V. (1984). Depth of search: How far can the retarded search through an internally represented problem space? In P. H. Brooks, R. Sperker, & C. McCauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 333-358). Hillsdale: Erlbaum.
- Tower of Hanoi. (n.d.). In *Wikipedia*. Retrieved from [http://en.wikipedia.org/wiki/Tower\\_of\\_Hanoi](http://en.wikipedia.org/wiki/Tower_of_Hanoi)
- White, S., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal of Abnormal Psychology*, *103*, 192-205.
- Wilson, A. C., Lengua, L. J., Tininenko, J., Taylor, A., & Trancik, A. (2009). Physiological profiles during delay of gratification: Associations with emotionality, self-regulation, and adjustment problems. *Journal of Applied Developmental Psychology*, *30*(6), 780-790. doi: 10.1016/j.appdev.2009.05.002
- Wright, R. A. (2008). Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Social and Personality Psychology Compass: Motivation and Emotion*, *2*, 682-701.

Wulfert, E., Block, J. A., Ana, E. S., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality, 70*(4), 533-552. doi: 10.1111/1467-6494.0501

Zelazo, P., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood: I. The development of executive function. *Monographs Of The Society For Research In Child Development, 68*(3), 11-27.

## Appendix

DCM was designed for children (Borys, 1984) and the games were shown by Klahr (1985) as suitable for the four to six-year-old participants. A common tactic of creating a harder game that is more age-appropriate for older participants is by adding a new game piece. Tower of Hanoi (ToH), for instance, has many versions from the two-disk version to the legendary 64-disk version. Borys, Spitz and Dorans (1982) studied the effect of ToH versions in children between six and 11 years of age. The older group of children, eight to 11-year olds, performed near ceiling with two-disk problems. However, when faced with three-disk problems, even the 11-year olds could only reach 30% of the maximum score.

The reason for the increased difficulty levels from the two-disk to the three-disk ToH lies in the state diagrams. In the two-disk version of ToH, there are only nine possible ways to arrange two disks on three pegs and there are 12 legal moves to translate among these states. In the three-disk version of ToH however, the possible arrangements and legal moves are higher, at 27 and 39 respectively. The number of permutations is the reason why the three-disk version allows for longer games than the two-disk version (Spitz & Borys, 1984). The increased degrees of freedom can be seen in the graphical representations of state diagrams where the nodes represent distributions of disks and the edges represent moves (Tower of Hanoi, n.d.). On the two-disk state diagram, the two furthest states are three steps apart; whereas on the three-disk state diagram, the furthest states are separated by seven edges. Meaning, the longest two-disk ToH problem takes three steps to solve whereas the longest three-disk ToH problem is seven steps in length.

The same principle applies to the differences between DCM and DCMR. The state diagram of DCM is relatively limited. The states can be viewed as a set of four values {A, B, C, and Empty}. There are  $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$  distinct permutations of these values. There are exactly 30 edges connecting the states to indicate legal moves. The state diagram of DCM was published in Figure 3 of Klahr (1985) and it allows for problems as long as eight steps.

To increase game difficulty, a more advanced class of DCM-type problems was designed; a game piece has been added and the degrees of freedom of movement increased (Cheung, 2007). The new four-object, five-vertex version, called DCMR, has an expanded state space (Figure 1). There are a total of  $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$  ways to arrange a set of four items on a network of five nodes {A, B, C, D, and Empty}. The state space of DCMR contains 120 states and 168 edges.

In order to locate the longest game allowed by this state diagram, a tool using the Java programming language was designed to process a state diagram via the breadth-first search algorithm (Blunden, 2005). The tool's output contained a listing of every possible solution path, ordered by length. According to this tool, the longest optimum solution was 12 steps in length.

By identifying DCMR's state diagram, the present study offers a master key to all DCMR problems and a design tool to locate a DCMR problem with desired game characteristics. The following procedures may be followed to find a game with hill-climbing opportunities (or goal-subgoal conflicts):

- a) Choose a pair of starting and ending states that are separated by the desired optimum solution.
- b) Examine whether the particular pair of states tempt participants to hill-climb at some point along the solution pathway (draw the states in trapezoids for easy visualization).
- c) If not, go back to step 1. If yes, compare the pathways with and without hill-climbing moves and make sure the former is longer than the latter (i.e. the optimum solution).

- d) If not, go back to step 1. If yes, stop and record the pair of states as the beginning and ending states. Place animals according to the beginning state and place foods according to the ending state.

After choosing the games, three implementations are recommended in order of ascending costs. The most economical way is to enlarge and print each game on a piece of paper (without the animals). Animal figurines from toy stores would suffice as game pieces. Game rules should be articulated first (Klahr, 1985) before letting participants move the animals on the paper. Researchers may either record participants on video for later coding, or directly observe the participants and code in real time. The coding of each game involves counting the total number of moves and a check box for hill-climbing, and the intercoder reliability for counts should easily reach 100%. Besides the printed version, costlier implementations of DCM(R) are to hire a woodshop to construct the board game or to work with a computer programmer on a virtual version. The advantage of these costlier versions is that the game pieces will be unable to move outside the grooves which will reduce the reliance on clear instructions at the onset. In the case of computers, coding should be built-in for instant and automatic results.