



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

심리학박사 학위논문

스트레스가 의사결정에
미치는 영향

2015년 2월

서울대학교 대학원

심리학과 임상신경심리 전공

박혜연

스트레스가 의사결정에 미치는 영향

지도 교수 최진영

이 논문을 심리학박사 학위논문으로 제출함
2014년 11월

서울대학교 대학원
심리학과 임상신경심리 전공
박혜연

박혜연의 심리학박사 학위논문을 인준함
2015년 1월

위원장 고성룡 (인)

부위원장 오성주 (인)

위원 최진영 (인)

위원 이대열 (인)

위원 김학진 (인)

초 록

무엇을 결정하거나 선택할 때, 두 종류의 인지신경학적 조절 시스템 간 경쟁적인 활동에 의해 행동이 결정된다는 가정이 일반적이다. 하나는 즉각적인 강화 여부에 따라 행동을 결정하는 ‘습관적 habit 혹은 모형 부재 model-free 행동 시스템’이며, 다른 하나는 행위자의 내적 상태나 외적 환경에 대한 지식 및 정보를 적극적으로 이용하여 행동을 결정하는 ‘목표지향적 goal-directed 혹은 모형 기반 model-based 행동 시스템’이다. 스트레스는 목표지향적인 행동을 방해하고 습관적 행동을 촉진하는 것으로 밝혀져 왔으며, 이는 스트레스가 두 행동 조절 시스템의 경쟁적 활동 과정에 개입할 가능성을 시사한다. 그러나 스트레스가 행동 선택 및 학습의 여러 요인에 미치는 구체적인 기전에 대해서는 아직 체계적 연구가 부족하다.

본 연구에서는 2 개 연구를 통해, 스트레스가 행동 선택의 과정 및 결과에 미치는 다면적인 영향에 대해 면밀하게 탐색했다. 연구 1에서는 습관적 행동 처리과정과 목표지향적 행동 처리과정을 구분하는 의사결정 과제, 2 단계 반전학습 과제를 개발하여, 실험실에서 유발된 급성 스트레스가 이 두 처리과정에 어떻게 관여하는지를 탐색했다. 정상 대학생들 스트레스 처치 조건과 비교통제 조건에 무선 할당했고, 피험자들의 과제 수행 행동에

강화학습의 계산모형을 적용하여 모형 기반 (model-based), 모형 부재 (model-free) 행동 경향성과 학습률 (learning rate) 을 추정했다. 실험 조건 간 과제 수행 행동 및 강화학습 모형 모수 추정치들을 비교한 결과, 스트레스 처치 조건에서는 비교통제 조건에 비해 모형 기반 (model-based) 행동이 저조했고, 강화 없는 상황에서의 모형 부재 (model-free) 행동 경향이 높았으며, 행동 선택 시 새로운 정보를 반영하는 경향, 즉 학습률 (learning rate) 이 저조했다.

이어서 연구 2 에서는 기능적 자기공명 뇌 영상 기법을 사용하여 의사결정에 대한 스트레스 효과를 신경활동 수준에서 탐색했으며, 스트레스의 행동에 대한 영향이 처치 정도에 따라 일관적인지 여키스-도슨 법칙 (Yerkes-Dodson law) 을 따르는지 확인했다. 정상 성인들을 스트레스 무처치 조건, 스트레스 단일처치 조건과 스트레스 이중처치 조건에 무선 할당하였고, 2 단계 반전 학습 과제를 수행하는 동안 기능적 뇌 영상을 촬영했다. 조건 간 과제 수행 행동 및 강화학습 모형 모수 추정치들을 비교한 결과, 스트레스 단일처치 조건의 피험자들은 무처치 조건에 비해 모형 기반의 (model-based) 목표지향적 행동이 증가하고 모형 부재의 (model-free) 행동 경향이 감소되었으나, 스트레스 수준이 더 높은 스트레스 이중처치 조건에서는 단일처치 조건에 비해 모형 기반 (model-based) 행동이 저조하게 나타났다. 스트레스와 관련된 인지행동의 양방향적 변화는 뇌신경 활동 수준에서도 확인되었다. 즉, 의사결정 시 내측 전전두엽 (medial prefrontal

cortex, 상측 측두엽 superior temporal cortex 의 신경활성화가 스트레스 처치 수준에 따라 증진되거나 저하되는 것으로 확인되었다. 이 두 영역의 신경활성화는 모형 기반 model-based 행동 경향을 반영하는 모수 추정치와 정적 상관관계인 것으로 나타났으며, 특히 내측 전전두엽 medial prefrontal cortex 의 의사결정 관련 신경활성화는 습관적 행동 지표와는 부적 상관관계였다. 또한 스트레스 처치는 우측 해마 hippocampus 의 선택 행동의 기대치 chosen value 관련 신경활성화를 저하시켰으며, 이는 행동적으로는 반전학습 reversal learning 수행 저하로 나타났다.

본 연구는 스트레스가 의사결정 시 습관적 행동을 증진시키는 인지행동적 기전을 밝히는 동시에, 스트레스의 효과가 그 정도에 따라 행동 선택의 여러 인지신경학적 요인에 다면적인 영향을 미침을 확인했다. 본 연구의 결과는 스트레스와 관련된 중독 행동 및 강박 행동 등 부적응적 행동의 병리적 기전 및 개입 방법에 대한 임상적 함의를 갖는다.

주요어: 급성 스트레스, 강화학습, 반전학습, 계산 모형, 기능적 자기공명 영상
학 번: 2007-30052

목 차

국문 초록.....	i
I. 서 론	1
1. 스트레스와 스트레스 반응	2
2. 스트레스와 의사결정	7
3. 의사결정에 대한 계산 모형.....	11
4. 연구 목적	16
II. 연 구 1	19
1. 연구방법	22
2. 연구결과.....	40
3. 논 의 1	52
III. 연 구 2	59
1. 연구방법	60
2. 연구결과.....	74
3. 논 의 2	95
IV. 종합 논의	102
참고문헌	106
영문 초록.....	119

표 목차

[표 1-1]	모형 비교 지표 값의 평균 (표준 오차).....	38
[표 1-2]	과제수행 결과의 조건 별 평균과 조건 간 차이.....	40
[표 1-3]	강화학습 모형의 조건 별 모수 추정치 사분위 값들.....	48
[표 1-4]	RL model 3의 델타 추정치들에 대한 회귀분석 결과.....	51
[표 2-1]	피험자들의 조건 별 과제수행 결과 평균(표준오차).....	74
[표 2-2]	RL model 2 모수 최대우도 추정치들의 조건 별 사분위 값들.....	81
[표 2-3]	모수들의 최대우도 추정치 평균들과 행동추정치들 간의 상관.....	81
[표 2-4]	뇌영상 분석을 위한 조건 별 모수추정 결과.....	89
[표 2-5]	실험조건 별 예측오류 관련 신경활성화 영역: 한 집단 t-검증.....	94

그림 목차

[그림 1-1]	두 단계 Markov 의사결정 과제의 구조.....	19
[그림 1-2]	두 단계 반전학습 과제.....	26
[그림 1-3]	반전학습에 대한 스트레스의 영향.....	42
[그림 1-4]	행동 선택 유지-전환 양상 분석.....	44
[그림 2-1]	fMRI 실험과제 중 한 시행의 타임라인.....	64
[그림 2-2]	스트레스 이중처치 조건에 대한 실험절차.....	67
[그림 2-3]	반전학습에 대한 스트레스 처치 수준 별 효과.....	77
[그림 2-4]	세 조건 별 행동 선택 유지-전환 양상 분석.....	79
[그림 2-5]	모형 기반 행동에 대한 스트레스 처치 수준 별 효과.....	83
[그림 2-6]	첫번째 의사결정 시 스트레스 단일처치로 인해 활성화가 증가된 영역.....	84
[그림 2-7]	의사결정 관련 신경활성화에 대한 스트레스의 양방향 효과.....	86
[그림 2-8]	선택 행동의 기대치 관련 BOLD 신호가 유의하게 나타난 영역...	90
[그림 2-9]	선택 행동의 기대치 관련 신경활성화가 나타난 영역 중 실험 조건 간 차이가 유의한 영역.....	91
[그림 2-10]	실험조건 별 예측오류 관련 BOLD 신호가 유의하게 나타난 영역.....	93

I. 서 론

우리는 늘 선택을 하며 살아간다. 점심에 무엇을 먹을까, 저녁에 운동을 할까 술을 마실까 등의 일상적인 선택에서부터 주식 매입/매각, 주택 대출금 지불 방식, 사고 현장에서의 신속한 대처와 같이 그 결과가 치명적인 경제적 손익이나 생명 유지와 직결되는 중요한 선택까지, 삶은 의사결정의 연속이다. 무엇을 선택해야 하는 상황에서 어떤 의사결정^{decision making}을 하는지는 주어진 옵션^{option}들에 대한 가치 판단에 따라 결정되나, 때로는 다른 변수가 의사결정에 결정적인 영향을 미치기도 한다. 그 중 스트레스는 행동 선택^{action selection}을 조절하는 주요 변수로, 스트레스가 우리의 행동에 영향을 미치는 예는 개인적인 기억을 통해서도 쉽게 찾을 수 있다. 중요한 시험이나 발표, 시합 등을 앞두고 우리는 그 결전의 날에 실행해야 하는 말이나 행동을 여러 차례 반복하여 연습한다. 매우 긴장되는 상황에서는 복잡한 사고 과정을 통한 문제 해결이나 순발력 있는 판단보다는 몸에 익은 습관으로 인해 자동적으로 행동하는 것이 더 쉽다는 것을 많은 경험을 통해 잘 알고 있기 때문이다. 또한 심리적 압박감에 시달리거나 불안한 상황에 처한 사람들은 오랫동안 애써서 유지해온 식이조절을 포기하고 기름지고 단 음식을 먹거나, 강한 의지력으로 실천하던 금연을 실패하고 담배를 다시 찾거나, 평소보다 많은 양의 술을 마신다. 정신 · 신체적으로 문제가 없는 정상 성인들도 스트레스 상황에서는 흡연 및 음주 행동이

증가하고 (Higgins and Marlatt, 1975; Hull and Young, 1983; Pomerleau and Pomerleau, 1987) 음식을 먹고 마시는 행동 역시 스트레스의 강도 및 특성에 따라 증가한다는 것 (Adam and Epel, 2007)은 많은 선행연구들에서 확인되었다. 즉, 스트레스는 습관적인 행동이나 감각추구 행동을 촉진하며, 행동의 나쁜 습관화라고 할 수 있는 중독의 발병 및 재발 과정에 중요한 위험 요소이다 (Le et al., 2005; Koob and Kreek, 2007; Cleck and Blendy, 2008; Sinha, 2008).

스트레스 상황에서 우리는 왜 인지적 노력을 기울여서 목표지향적으로 행동하기 보다는, 부정적인 결과를 초래할지도 모르는 (혹은 그럴 것이 분명한) 행동을 하며 즉각적 만족을 추구하거나 과거 습관에 의지하게 되는 것일까? 이 질문의 답을 얻기 위해서는 스트레스가 의사결정 행동의 인지적 과정에 미치는 영향 및 의사결정에 관여하는 뇌 구조물 및 그 기능이 스트레스 반응으로 인해 어떻게 변화하는 지를 탐색해야 하며, 이는 사회경제적이나 국민정신건강 측면에서 매우 의미 있는 일이다.

1. 스트레스와 스트레스 반응

스트레스_{stress}는 본래 물체에 가해지는 물리적인 힘을 뜻하는 용어이다. 이 용어를 생물체를 위협하는 자극 또는 그 자극으로 인한

생리적 반응을 가리키는 용어로 쓰기 시작한 것은 캐나다의 생리학자이자 의사였던 한스 셀리에(Hans Selye)로, 그는 신체에 부담을 주는 압박감으로 인하여 신체에 비특정적인 반응이 일어나는 것이라고 스트레스를 정의했다(Selye, 1976). 그러나 유기체의 생명을 위협하는 자극과 그로 인한 생리적 반응이 스트레스라는 용어로 설명되기 이전에 이미 월터 캐논(Walter Cannon)이라는 생리학자가 동물이 생명의 위협에 직면했을 때 일련의 생리적 반응을 통해 ‘싸우기 또는 도망가기(fight or flight)’를 한다(Cannon, 1929)고 주장한 바 있듯이, 유기체의 신체적 항상성을 깨뜨리는 위협적인 자극과 이에 대한 신체 반응은 생리학 및 의학에 있어서 오랜 관심사였다.

현재 스트레스는 보다 광의의 개념, ‘신체의 신항상성(allostasis)을 깨뜨리는 모든 것’으로 정의되며, 스트레스로 인한 신체의 반응은 ‘스트레스 상황에서 신체가 신항상성(allostasis)을 되찾으려 하는 시도’로 통용되고 있다(Sapolsky, 1996). 이때 신항상성(allostasis)이란 신체가 외부환경의 변화에 적응하기 위해 스스로 역동적으로 변화함으로써 항상성(homeostasis)을 찾아가는 과정(Flier et al., 1998; McEwen, 2000)을 뜻한다. 또한 스트레스 반응을 일으키는 요인(stressor)에 대한 정의도 광범위해져서, 전통적인 스트레스 연구에서는 스트레스 요인을 생명을 위협하는 신체적 위기에 국한시켰으나 현대심리학은 심리적인 문제나 사회적 혼란 역시 스트레스 반응을 유발한다는 것을 밝혔다.

유기체가 스트레스 상황에 처하면 스트레스 자극이 무엇이건 간에

유기체 내부에서 일련의 생리적 반응이 일어난다(Ulrich-Lai and Herman, 2009). 교감신경계(sympathetic nervous system)에서 에피네프린(epinephrine)과 노르에피네프린(norepinephrine)이 분비되는 한편, 뇌의 시상하부(hypothalamus)와 뇌하수체(pituitary)에서는 각각 부신피질 자극 호르몬 방출 호르몬(corticotrophin releasing hormone, CRH)과 부신피질 자극 호르몬(adrenocorticotrophic hormone, ACTH)이 순차적으로 분비되어 그 결과로 부신피질(adrenal cortex)에서 스테로이드 호르몬인 글루코코르티코이드(glucocorticoid)가 분비된다(O'Connor et al., 2000). 이러한 생리적 스트레스 반응들로 인해 신체는 소화, 성장, 생식 작용이 억제되고, 포도당, 단백질, 지방 등의 에너지 요소가 저장되는 대신 근육으로 이동하며, 심박수, 혈압, 호흡량이 증가한다. 일상 생활에서 섭식을 통해 생활에 필요한 에너지를 얻고 생리적, 행동적 일관성을 유지하며 살아가던 유기체가 일상적인 상태에서 벗어나 스트레스에 대처하게 되는 것이다. 스트레스 반응들은 스트레스 유발 원인이 사라지면 사라지고, 유기체는 다시 일상적인 안정 상태로 돌아간다.

스트레스 반응은 생존에 도움이 되고 적응적인 측면이 있다. 스트레스 상황에서는 유기체가 전투 체계를 갖춰서 외부의 위협에 맞서 싸우는 데에 적합해지며, 일시적인 긴장감이나 심리적 압박감은 각성 수준을 상승시켜 정신적 활동이 오히려 활발해지고 생산성이 증진되는 효과가 있다(Anderson, 1976; Cahill and Alkire, 2003). 즉 경미한 스트레스 상황에서는 교감신경계가 활성화 되고 체내

카테콜아민^{catecholamine}이나 글루코코르티코이드^{glucocorticoid} 수준이 높아지면서 감각 자극에 대한 민감성이 높아지고 주의력^{attention}, 기억력^{memory} 등 인지 기능이 증진된다. 그러나 셀리에^{Hans Selye}가 스트레스 반응을 세 단계로 구분하여 마지막 단계를 스트레스가 만성적으로 유지될 때 질병이 생기는 단계로 정의했듯이(첫 번째 단계는 스트레스를 인식하는 단계, 두 번째 단계는 스트레스 반응 체계가 활성화되어 신항상성^{allostasis}이 변화하는 단계로 정의했음)(Selye, 1946), 스트레스가 심하거나 오랫동안 지속되면 여러 가지 문제가 발생한다. 심혈관계 질환 및 대사 증후군 발병 가능성이 높아지고 면역계 기능이 억제되며 수면과 섭식 등 기본적인 생명 유지 기능이 저하될 뿐 아니라, 중추신경계에서도 기능적 및 구조적 손상이 나타난다(Sapolsky, 1996; De Kloet et al., 2005; Schneiderman et al., 2005).

중추신경계의 대표적인 스트레스 반응 체계인 시상하부-뇌하수체-부신피질^{hypothalamus-pituitary-adrenal cortex}의 네트워크는 대뇌의 전두엽과 해마와 같은 구조물들로부터 부적 피드백을 받아 체내 글루코코르티코이드^{glucocorticoid} 수준을 조절한다. 그러나 지속적으로 심한 스트레스에 노출되면, 전두엽, 해마 등 스테로이드 호르몬의 수용체가 밀집되어 있는 구조물들에서 구조적, 기능적 변화가 나타나, 스트레스 호르몬의 분비 체계의 균형이 깨지게 되고 이는 악순환으로 연결된다(De Kloet et al., 2005).

해마^{hippocampus}의 경우 스트레스 반응이 일어나면 CA1에서

장기강화작용(long term potentiation)이 감소하며, 만성적인 스트레스에 노출될 경우 CA3 영역의 피라미드 세포(pyramidal cell)의 수상돌기 위축 현상 및 세포사(cell death)가 일어난다(Kim and Diamond, 2002). 전전두엽(prefrontal cortex) 역시 만성적 스트레스에 취약하여, 내측 전전두엽의 상층(layers 2-3) 피라미드 뉴런들은 만성적 스트레스로 인해 수상돌기 위축, 재배열이 일어나며, 수상돌기 가지(dendritic spine)가 감소하여 흥분성 정보 입력이 상당히 감소하는 것으로 확인되었다(McEwen and Sapolsky, 1995). 편도체(amygdala)는 반복적인 스트레스에 노출될 경우, 다른 구조물들과 달리 구조적인 팽대(hypertrophy) 현상이 나타난다. 즉 편도체(amygdala)의 기저 외측핵(basolateral nucleus)에 있는 신경세포들(pyramidal-like and stellate neurons)의 수상돌기(dendrite) 길이가 길어지고 신경세포 가지들이 뻗어나가는 양상도 증가한다(Mitra and Sapolsky, 2008; Roozendaal et al., 2009).

대뇌 주요 구조물들에서 일어나는 스트레스로 인한 구조적 변화는 기능적 변화와 동반되어, 기억력, 주의력 등 인지행동에도 영향을 미친다. 지속적인 스트레스에 노출되면 해마의 구조적, 기능적 변화로 인해 일화 기억(episodic memory)이 저하되며, 스트레스 상황에서는 전전두엽(prefrontal cortex) 기능인 작업 기억(working memory)이 저하되고 선택적 주의력(selective attention)이 감소하는 것으로 밝혀졌다(Murphy et al., 1996; Mizoguchi et al., 2000; Liston et al., 2006). 스트레스에 의해 구조적, 기능적 변화를 겪는 해마(hippocampus), 전전두엽(prefrontal cortex) 등의 대뇌 구조물들은 기억력(memory), 주의력(attention) 등의 인지 기능뿐 아니라

의사결정 과정 및 결과에도 영향을 미칠 것이 예상되나, 아직 이에 대해 밝혀진 것은 적다.

2. 스트레스와 의사결정

의사결정^{decision making}이란 각기 다른 결과를 초래하는 것으로 예상되는 여러 옵션^{option}들 중 하나를 고르거나 선택 여부를 결정하는 인지행동적 과정을 일컫는 추상적인 개념이다. 전통적인 행동주의적 관점에서는 행동 선택이 이루어지는 원리에 대해서는 두 가지 가설이 있다. 하나는 유기체는 행동의 즉각적인 결과에 기반하여 행동을 선택하며(Thorndike, 1911), 행동 선택의 결정적인 요인은 그 결과로 자신의 추동^{drive}이 감소되는지 여부라는 것이다(Hull, 1943; Spence, 1956). 즉 자신이 선택한 행동의 즉각적인 결과로 욕구가 해결될 때 그 행동이 강화된다는^{reinforced} 것인데, 이러한 관점에서 반복적인 강화 경험에 따라 행동이 조형되어 습관화되므로, 주어진 자극과 그에 대한 행동(선택 행동 혹은 반응)을 보상 받은 횟수의 함수로 가정할 수 있다. 반면 행동 선택의 원리에 대한 다른 가설은 행동이 목표지향적으로 이루어진다는 것이다(Tolman, 1951). 즉 행동이 즉각적으로 보상받지(강화되지) 않는다고 해도 유기체는 자신의 내적 상태나 외적 상황을 고려하여 앞으로의 결과를 예상하고 평가 및 비교하는 과정을

통해서 행동을 선택한다는 것이다. 전자는 인지신경학적으로 습관 체계(habit system), 후자는 목표지향 체계(goal-directed system)로 구분되며, 신경해부학적 실체가 밝혀져 왔다(Doll et al., 2012; Penner and Mizumori, 2012; Lee, 2013). 의사결정 행동의 원리에 대한 오래된 이 논쟁에 대해, 현재는 두 체계 간의 경쟁적인 행동 조절의 결과로 행동이 결정된다는 합의가 이루어졌으며, 둘이 서로 균형을 이루는 데에 있어서 어떤 관계를 맺고 있는지, 두 조절 과정을 변화시키는 조절 요인이 있는지 등에 대한 연구가 지속되고 있다.

행위자(agent)의 행동 선택이 습관 체계에만 의존하는지, 다시 말하면 전적으로 즉각적인 결과에 따라서만 행동 선택이 이루어지는지, 아니면 내적 상태나, 외적 상황에 대한 정보를 이용하여 목표지향적으로 이루어지는지를 탐색하는 데에는 전통적으로 ‘수반성 감소(contingency degradation)’나 ‘결과 가치 감소(outcome devaluation)’ 등의 행동 조형 실험 패러다임이 사용되어왔다(Dickinson, 1985; Balleine and Dickinson, 1998, 2005). ‘수반성 감소(contingency degradation)’ 패러다임에서는 동물에게 특정 행동을 보상을 통해 학습시킨 후, 그 행동과 결과 간의 수반성(contingency)을 저하시켜서 이후 행동 변화를 관찰한다. 예를 들면 쥐에게 레버를 눌러서 단물을 섭취하게 학습시킨 후, 레버를 누르지 않아도 간헐적으로 단물을 먹을 수 있는 절차를 거치게 하여, 레버를 누르는 행동과 단물이라는 결과 간의 수반성(contingency)을 감소시키는 것이다. 반면, ‘결과 가치 감소(outcome devaluation)’ 패러다임에서는 특정

행동을 강화시킨 후, 그 행동의 결과 가치를 저하시킨다. 예를 들어, 쥐들에게 레버를 눌러서 단물 섭취하는 것을 학습하게 한 후, 단물을 실컷 먹게 하여 포만감을 느끼게 한다. 포만감을 느끼게 된 쥐들에게서는 단물에 대한 내적 가치가 감소된다. 이 두 실험 패러다임을 적용한 연구 결과, 동물들은 재평가 절차를 거친 후 결과와의 수반성(contingency)이 감소된 행동이나 결과 가치가 감소된 행동을 선택하는 빈도가 감소하는 것으로 나타났으며, 이는 동물들이 행동의 결과에만 의존적으로 행동 선택하는 것이 아니라 자신의 내적, 외적 상황에 대한 심리적 모형을 가지고 목표지향적 행동을 한다는 증거가 된다(Doll et al., 2012).

그러나 스트레스 처치를 받은 동물은 재평가(revaluation) 패러다임을 이용한 스트레스 실험 연구에서, 행동과 그 결과 간 수반성(contingency)이 저하되거나 그 행동의 가치가 감소한 경우에도 이전 학습 행동의 빈도가 줄어들지 않았으며(Dias-Ferreira et al., 2009), 이는 정상 대학생에게서도 확인되었다(Schwabe and Wolf, 2009). 특히 동물 연구 결과에서는 스트레스 처치 받은 쥐들에게서 습관적인 행동 양상이 대뇌의 구조적 변화와 수반되었다. 즉 스트레스 군의 쥐들은 통제군에 비해, 목표지향적 행동의 주요 영역으로 알려진 전두엽 일부 영역과 등외측 선조체(dorsolateral striatum)의 부피가 축소되어 있었고, 습관 형성과 관련되어 있는 것으로 알려진 등내측 선조체(dorsomedial striatum)의 부피는 증가되어 있었다(Dias-Ferreira et al., 2009).

한편 스트레스 처치 받은 동물들은 미로 찾기 실험 중에도 인지적인 기억력을 기반으로 행동하기 보다는 자극에 즉각적으로 반응하는 방식으로 행동하는 양상을 보였으며(Kim and Diamond, 2002; Henckens et al., 2009; Ferragud et al., 2010), 이는 스트레스가 감각 처리를 증진시키며 인지적인 기억력에 기반하는 행동 대신 자극에 단순히 반응하는 방식으로 행동하는 양상을 증진시킨다는 가설의 증거가 된다.

스트레스가 인지 기능이 요구되는 지적 과제 수행 중 행위자_{agent}의 감각추구 행동을 촉진시키고, 환경적 변화에 대해서 이전 학습 행동을 반복하는 경향을 증진시킨다는 것이 확인되었으나, 행동 선택에 대한 스트레스 효과에 대해서는 몇 가지 의문점이 남아 있다. 첫 번째로, 전통적인 재평가_{revaluation} 실험 패러다임을 이용한 연구에서는 스트레스가 섭식행동에 미치는 영향이 통제되지 못했다는 문제가 있다. 스트레스는 식욕을 증가시키며, 스트레스 호르몬인 글루코코르티코이드_{glucocorticoid}는 식욕감소호르몬에 대한 뇌의 민감성을 저하시키는 것으로 알려져 있다(Adam and Epel, 2007). 따라서 스트레스가 행동선택에 미치는 영향을 탐색하기 위해서는 음식 이외에 다른 보상이 제공되는 행동에 대해서도 스트레스의 영향이 일관적으로 나타나는지 확인할 필요가 있으며, 이를 위한 새로운 실험 패러다임의 적용이 필요하다. 두 번째로, 스트레스가 습관 행동을 촉진한 것이 목표지향 체계의 행동 과정을 방해함으로써 일어난 보상적인 현상인지, 즉각적인 결과에 의존적인

습관 체계의 행동 과정이 독립적으로 촉진된 것에 기인하는지 명확하지 않다. 이와 관련해서 스트레스가 목표지향적 행동을 방해하고 습관 행동을 촉진한다기 보다는 환경적 변화에 대한 유연한 대처를 방해하는 것은 아닌지에 대한 의문점도 제기된다. 세 번째로, 스트레스가 행동 선택에 미치는 영향에 기저하는 뇌신경 활동 양상의 변화가 아직 명확하지 않으므로, 스트레스가 습관행동을 촉진하고 목표지향적 행동을 방해하는 데에 어떤 뇌 영역의 활성화 변화 양상이 연관되어 있는지를 밝히는 것이 필요하다. 마지막으로, 선행 연구에서는 스트레스가 습관 행동을 촉진시키는 것으로 보고되었지만, 행동 선택에 대한 스트레스의 영향이 항상 같은 방식으로 나타나는 것인지 알 수 없으며, 스트레스의 정도에 따라 양방향성의 영향을 미칠 가능성을 확인해야 한다.

3. 의사결정에 대한 계산 모형

최근에는 의사결정 행동에 대한 목표지향적 체계(goal-directed system)와 습관적 체계(habit system)의 역할을 탐색하는 데에, 행동과 그 결과 간의 관계를 직접 조작하여 목표지향적 행동과 습관행동을 구분하는 전통적인 실험 기법 대신, 의사결정자의 행동 선택에 ‘강화학습(reinforcement learning (Sutton and Barto, 1998))’이라 불리는 계산 모형(computational model)을 적용하는 기계학습식(machine learning) 접근 방식이 활용되고 있다. 강화학습의 계산 모형에서는 행동에 대한 가치 함수(value function)가 설정되며, 가치

함수는 해당 상황에서 특정 행동(옵션 선택 행동)의 가치가 의사결정자의 주관적인 추정으로 결정된다. 의사결정자가 상황에 대한 지식이나 가설 없이 결과에 기반하여 행동을 선택하는 강화학습의 수학적 모형은 오래 전부터 기계학습(machine learning) 및 신경과학 분야에서 활용되어 왔다. 가장 기본적인 강화학습 모형은 고전적 조건형성의 수학적 알고리즘인 Rescorla-Wagner 모형이다(Rescorla and Wagner, 1972; Sutton and Barto, 1998) (식 (1)). 이 모형에서는 행위자(agent)가 매번 자신이 선택할 수 있는 옵션(option)에 대해 기대치(expected value)를 계산하여 행동을 선택하며, 옵션의 기대치는 실제 가치(그 옵션을 선택한 행동의 결과로 경험하는)와의 차이, 즉 보상에측오류(reward prediction error)에 의해 업데이트된다.

$$V_{t+1}(x) = V_t(x) + \alpha [R_t(x) - V_t(x)] \quad (1)$$

식 (1)에서 $V_t(x)$ 는 t 번째 시행에서 옵션 x 를 선택했을 때, x 의 기대치에 대한 함수, 즉 가치 함수이다. $t + 1$ 번째 시행에서의 x 의 기대치 $V_{t+1}(x)$ 는 그 시행에서 x 의 실제 가치 $R_t(x)$ 와의 차이, 즉 보상에측오류(reward prediction error) $(R_t(x) - V_t(x))$ 에 의해 수정되며, 이때 보상에측오류는 의사결정자가 새로운 정보를 다음 행동에 반영하는 정도인 학습률(learning rate) α 에 의해 조정된다.

Sutton과 Barto는 Rescorla-Wagner 모형에 사건의 시간차, 즉

발생 시점에 따른 경험의 차등적인 영향을 반영시켜, 시간차 학습(temporal difference learning, TD) 모형을 개발했다(Sutton and Barto, 1998). TD 모형은 행동-결과의 경험이 시간에 따라 약화되어 이후 행동에 반영된다. 현재 상태를 S_t 이라 할 때, $V(S_t)$ 은 앞으로 그 상태를 통해 얻게 될 결과들의 합으로 표현 가능한다(식 (2)), 이 때 앞으로 얻게 될 결과들은 현재 시점에서 멀수록 절감되어 현재 상태에 영향을 미친다. 따라서 TD 모형에서 보상에측오류(reward prediction error)는 Rescolar-Wagner 모형과 달리, $R_t + \gamma V(S_{t+1}) - V(S_t)$ 의 형태가 된다(식 (3)).

$$V(S_t) = E[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \dots | S_t] \quad (2)$$

$$V_{new}(S_t) = V_{old}(S_t) + \alpha [R_t + \gamma V_{old}(S_{t+1}) - V_{old}(S_t)] \quad (3)$$

Rescolar-Wagner 강화학습 모형과 TD 학습 모형은, 행동 모형 알고리즘에 행위자의 내적 상태나 외부 환경에 대한 정보가 반영되지 않으므로, 행동이 오로지 시행 착오(trial-and-error)의 축적으로 조형되며 매번 즉각적인 결과에 기인한다는 측면에서 습관 체계의 행동 특성을 구현한다. 최근 의사결정 이론에서는 이를 상황에 대한 심적 표상(metal representation) 없이 이루어진다는 뜻에서 모형 부재(model-free) 강화학습 또는 의사결정이라고 지칭한다(Doll et al., 2012).

목표지향 체계의 행동 선택은 행동의 즉각적인 결과에 따라 이루어지기 보다는, 최고의 결과를 얻을 수 있는 행동을 선택하기 위해 행위자의 내적 상태 및 외부 상황을 고려한 결과 예측을 고려한다.

따라서 행위자는 의사결정을 둘러싼 제반 상황에 대해 심리적 수준의 표상 즉, 모형화(modeling)를 하게 된다. 목표지향적 행동 선택은 동적 계획법(dynamic programming)이라는 계산적 문제해결 방법을 활용하여 수학적 모형으로 구현된다(Sutton and Barto, 1998; Doll et al., 2012). 동적 계획법을 활용하는 강화학습 모형의 최적의 의사결정은 현재 상황에서 최고의 결과를 얻을 수 있는 최선의 행동을 하는 것이다.

$$V_{t+1}(\mathbf{x}) = \max V_t(\mathbf{x}) \quad (4)$$

즉 행위자는 자신의 상황에 대한 정보 및 지식을 갖고 있으며 이를 활용하여 최대의 이익을 계산한다는 가정을 하므로(식 (4)), 강화학습 가치함수 알고리즘에 상황에 대한 정보, 이를테면 행동 선택으로 인한 단계의 전환 및 보상 획득 확률(reward probability) 등이 반영된다. 이 경우, 행위자(agent)의 행동이 상황에 대한 지식을 기반으로 한 심적 모형(model)과 그로 인한 시뮬레이션으로 이루어지므로, 목표지향 체계의 강화학습 모형에서의 의사결정은 모형 기반(model-based)의 강화학습 또는 의사결정이라 지칭한다.

행위자의 의사결정 특성을 탐색하는 데에 있어서 강화학습 모형을 적용하는 연구 방법은 행동 자료에 대한 평균치에 근거한 통계 분석 방법과 달리 다음과 같은 장점이 있다. 첫째로, 강화학습에 대한 계산적 접근 시, 모형의 계산 식에 의사결정 행동의 가설적 요인을 반영하여 실제 수행 자료에 잘 부합하는 지 확인하는 것이 가능하고, 이론적

가설에 의해 모형에 포함된 모수 parameter 를 추정함으로써 행위자의 개인적 행동 특성을 양적으로 추정할 수 있다. 또한 강화학습 이론은 행동 선택에는 행위자가 이전에 경험한 모든 행동-결과가 반영되며 현 행동 역시 이후의 행동 선택에 지속적으로 영향력을 갖는다는 것을 가정한다. 따라서 강화학습의 계산 모형 기법을 통해, 행위자들의 과제 수행 자료, 즉 과제 내 모든 시행의 행동 선택과 결과의 축적 action-outcome history 를 통해 개인적 강화학습 특성을 탐색하는 것이 가능하다.

의사결정에 대한 계산적 강화학습 모형 기법은 뇌 영상 기법과 통합적으로 활용됨으로써, 강화학습의 이론적 가정들의 신경활동적 기반이 확인되어 왔다(O'Doherty et al., 2007). 특히 행동의 결과(보상 여부) 및 기대한 결과와 실제 결과 가치 간의 차이, 즉 보상 예측 오류reward prediction error는 도파민성 신경세포의 신경신호로 전달되며 배측 선조체ventral striatum의 신경활성화에 반영됨이 일찍이 밝혀진 바 있으며 (Schultz et al., 1997; Hollerman and Schultz, 1998; Schultz, 1998), 선택한 행동(혹은 자극)의 가치(예측되는 기대치expected value)가 내측 측두엽medial temporal cortex을 비롯한 여러 뇌 영역의 신경활동 신호에 반영되는 것이 확인되었다(Kim et al., 2006; Schönberg et al., 2007; Hare et al., 2008). 이 뿐 아니라, 최근에는 목표지향적 행동과 습관적 행동을 강화학습 모형에 구별적으로 반영하여 두 행동 체계의 행동 조절 과정을 비교하는 시도가 이루어지고 있는 등(Daw et al., 2011; Doll et al., 2012; Otto et al., 2013a; Lee et al., 2014), 인간의 선택 행동의

특성을 밝히는 데에 계산 모형 기법이 활발하게 사용되고 있다.

4. 연구 목적

본 연구는 강화학습의 이론적 틀에 근거하여, 스트레스가 의사결정의 다각적 요인에 미치는 영향을 탐색하고자 했다. 즉, 스트레스가 1) 습관적 행동 경향과 목표지향적 행동에 각각 어떤 영향을 미치는지, 2) 행동-결과 관계가 가변적인 상황에서 의사결정자가 변화 정보를 반영하는 정도에 어떤 영향을 미치는지 등을, 계산적 강화학습 모형 및 기능적 뇌 영상 실험을 통해 밝히고자 했다. 또한 스트레스와 의사결정 간의 관계가 스트레스의 정도에 따라 일관적으로 혹은 단조적으로^{monotonic} 나타나는지 아니면 양방향적^{bidirectional} 혹은 비단조적으로^{non-monotonic} 나타나는지를 확인하고자 했다. 이상의 연구 목적을 위해 두 단계의 연구를 구성하여 진행했으며, 스트레스 효과는 실험실에서 일시적으로 유발시키는 급성 스트레스^{acute stress}에 한정시켰다. 각 연구의 개요는 다음과 같다.

1) 연구 1: 스트레스로 인한 두 행동 조절 체계의 영향력 변화 탐색

연구 1에서는 스트레스로 인해 보상 기반 의사결정 행동 과정 중

모형 기반_{model-based} 행동 경향성 및 모형 부재_{model-free} 행동 경향성이 어떻게 변화하는지를 정상 성인들을 대상으로 탐색했다. 선행 실험 연구에서 효과적으로 스트레스 반응을 유발하는 것으로 확인된 스트레스 유발 실험을 스트레스 실험군에 처치한 후, 스트레스 실험군과 통제군의 의사결정 행동 특성을 비교 분석했다. 연구참여자들의 의사결정 행동을 측정하는 데에는 목표지향적 행동과 습관 행동을 구분할 수 있는 실험 과제(2 단계 반전 학습 의사결정 과제_{2-stage reversal learning decision task})를 개발하여 사용했고, 각 피험자의 의사결정 과제 수행 특성은 강화학습_{reinforcement learning}의 이론적 틀을 적용하여 분석했다.

2) 연구 2: 의사결정에 대한 스트레스의 정도에 따른 영향과 그 기저 신경활동적 특성

연구 2에서는 연구 1에서 탐색한 스트레스의 영향의 신경활동적 기반을 탐색함과 동시에, 스트레스의 영향이 처치 정도에 따라 달라지는지 여부를 확인하고 그와 관련된 신경활성화 양상을 탐색하고자 했다. 정상 성인들을 스트레스 무처치_{no-stress-treatment} 조건, 스트레스 단일처치_{single-stress-treatment} 조건, 스트레스 이중처치_{double-stress-treatment} 조건에 무선 할당하여, 스트레스 처치 유무와 정도에 따른 의사결정 양상을 비교 분석하였다. 연구는 기능적 자기공명영상 실험을 통해

스트레스 관련 의사결정 행동에 기저하는 대뇌 신경 활성화 양상을 확인했다.

II. 연구 1

스트레스가 행동 선택에 미치는 영향을 탐색하기 위해, 정상 대학생들의 의사결정 행동을 계산적 강화학습 모형 기법을 적용하여 분석했다. 의사결정 과제로는, 모형 기반 강화학습(model-based reinforcement learning) 행동과 모형 부재 강화학습(model-free reinforcement learning) 행동을 적절히 구분하는 것으로 밝혀진 ‘두 단계 Markov 의사결정 과제(2-stage Markov decision task (Daw et al., 2011)(그림 1-1)’를 변형하여 사용했다.

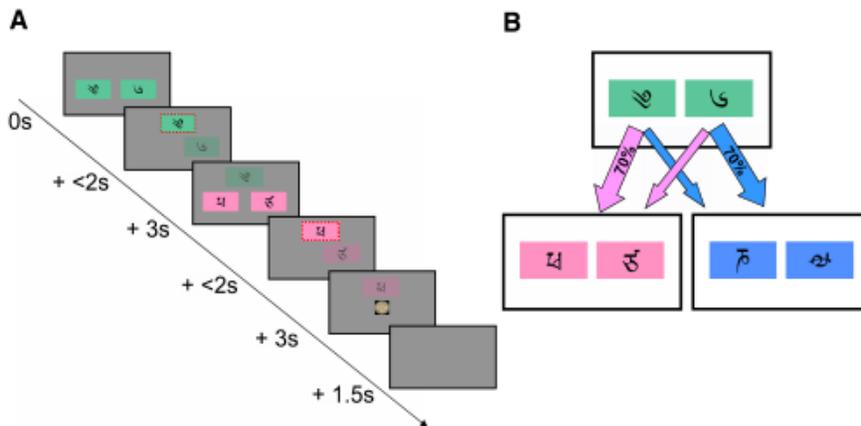


그림 1-1. 두 단계 Markov 의사결정 과제의 구조. (A) 시행 내의 사건 진행 타임라인. 첫 단계에서의 옵션 선택이 다음 단계로 이어지고, 두 번째 단계에서의 옵션 선택은 매 번 가우시안 랜덤 워크에 의해 바뀌는 보상 확률에 따라 1 파운드로 강화됨. (B) 단계전환구조. 첫 번째 단계의 선택은 두 번째 단계의 두 상태 (분홍 혹은 파란 화면) 중 하나와 더 빈번하게 연결됨(70%의 확률). (Daw et al., 2011)

선행 연구에서 개발된 두 단계 Markov 의사결정 과제_{2-stage Markov decision task}는 한 시행_{trial}이 두 단계_{stage}로 이루어져 있으며, 두 번째 단계의 상태_{state}가 첫 번째 단계의 행동 선택에 따라 결정되는 Markov 의사결정 과정¹을 따른다. 즉 행위자_{agent}는 과제의 매 시행마다 두 단계에 걸쳐 각각 한 번씩의 행동 선택을 해야 하는데, 각 시행에서 다음 단계의 상태 s' 는 현재 상태 s 와 현재 상태에서의 의사결정자의 행동 a 에 의해, 지정된 확률 $P_a(s, s')$ 에 따라 결정된다. 따라서 첫 번째 단계에서의 행동 선택은 다음 단계가 어떤 상태일지에 영향을 미칠 뿐 아니라, 다음 단계에서 하는 행동의 결과와도 서로 관련성이 있다.

이 과제에서는 첫 번째 단계에서 행동을 선택할 때 단계 전환 확률_{stage transition probability} $P_a(s, s')$ 을 고려해야 하므로, 단계 전환_{stage transition} 방식, 즉 과제의 전체적인 구조에 대해 심적 모형화_{modeling}가 필요하다. 따라서 첫 번째 단계에서 하는 선택 행동은 직전 시행_{previous trial}에서의 행동 보상 여부와 함께 단계 전환 확률_{stage transition probability}이 함께 고려되어야 하므로, 습관 체계의 모형 부재_{model-free} 행동 과정과 목표지향 체계의 모형 기반_{model-based} 행동 과정이 모두 기여하게 된다. 선행 연구에서는 이 과제의 첫 번째 단계 행동 선택 과정을 모형 기반

¹ Markov 의사결정 과정(Markov Decision Process, MDP)은 러시아의 수학자 안드레이 마코프(Andrey Markov)의 이름을 딴 의사결정 과정의 수학적 프레임 중 하나이다. MDP는 의사결정자의 컨트롤이 부분적으로만 영향을 미치는 상황에서의 각기 분리된 시간적 단계로 이루어지는 확률적 의사결정 과정을 모형화하는 방식이다.

Howard RA (1960) DYNAMIC PROGRAMMING AND MARKOV PROCESSES. In. USA: Massachusetts Institute of Technology Press.

및 모형 부재 행동 과정이 서로 선형적인 보완(linear trade-off)을 이루는 수학적 모형 식으로 구현된 바 있다(Daw et al., 2011).

본 연구에서는 두 단계 Markov 의사결정 과제(2-stage Markov decision task)에 반전 학습(reversal learning) 패러다임을 통합한 의사결정 과제를 개발하여 사용했다. 반전 학습 패러다임은 과제의 구간마다 단서 없이 행동과 결과 간 수반성(contingency)을 반전시키므로, 반전 학습 과제에서 행위자는 행동-결과 간 수반성(contingency) 변화에 따라 유연하게 행동을 변화시키는 것이 필요하다. 특히 선택 행동이 보상과 확률적으로 연결되어 있는 확률적 반전 학습 과제(probabilistic reversal learning task)에서는, 행위자가 자신의 행동 결과를 최대화하기 위해 매번 새로운 경험을 축적하고 이전의 경험과 통합하여 그것을 바탕으로 상황적 변화를 파악하고 유연하게 대처해야 한다. 따라서 반전 학습 패러다임이 통합된 본 과제는 ‘두 단계 Markov 의사결정 과제(2-stage Markov decision task)’에 비해 모형 기반(model-based) 의사결정 과정 및 일화기억(episodic memory)이 더욱 요구되는 특징이 있다.

정상 대학생들을 스트레스 조건과 통제 조건에 무선 할당한 후 의사결정 과제를 수행하게 했고, 과제 수행 결과로 목표지향적 혹은 습관적 행동 특성과 반전 학습(reversal learning)의 정도 등을 조건 간 비교 분석했다. 또한 피험자 별로 과제의 매 시행마다 축적된 행동-결과 기록(action-outcome history)에 모형 부재(model-free) 및 모형 기반(model-based) 행동 선택을 구현한 강화학습(reinforcement learning)의 계산 모형(computational model)을

적용했다. 그 결과 각 피험자들의 모형 부재_{model-free} 및 모형 기반_{model-based} 행동 경향, 새로운 정보를 행동에 반영하는 정도(학습률_{learning rate}) 등의 의사결정 특성을 양적인 수치로 추정했고, 이에 대한 스트레스 효과를 탐색했다.

1. 연구방법

1) 연구참여자

56명의 정상 성인(연령 평균 20.36세, 표준편차 1.91; 남자 27명, 여자 29명)들을 대상으로 했다. 서울대학교에서 심리학 관련 수업을 수강하는 학부생들 중 연구 참여를 희망했던 자원자들과 학교 인터넷 게시판에 게시한 연구참여자 모집 광고를 통해 신청한 학생들을 대상으로 참여 조건에 맞는 사람들을 선별하여 모집하였다. 선별 검사를 통해 의식을 잃을 만큼 심한 두부 외상이 없었던 자, 항정신성 약물이나 스테로이드 등 중추신경계에 영향을 미치는 약물을 복용하지 않는 자, 정신과 병력이 없는 자, 최근 심각한 스트레스 사건을 경험하지 않는 자를 선정했고, 스트레스 호르몬 반응에 영향을 미치는 것으로 밝혀진 담배 및 경구성 피임약 사용자 역시 제외시켰다(Kirschbaum et al., 1999; Mendelson et al., 2005). 심리사회적 스트레스에 대한

시상하부-뇌하수체-부신피질(hypothalamus-pituitary-adrenal cortex) 네트워크의 스트레스 반응에 성별 차이가 있는 것으로 밝혀졌으나, 생리 주기 중 황체기(luteal phase)에 해당되는 여성은 남성과 체액의 코티솔(cortisol) 반응에서 차이가 없다(Kirschbaum et al., 1999)는 연구 결과에 근거하여, 생리 주기의 21일 이후에 해당되는 여성은 남성과 함께 연구에 포함시켰다. 연구대상들은 각각 비교 통제 조건과 스트레스 실험 조건에 무선 할당되었고, 실험 전 6시간 동안은 자율신경계 각성이 유발되는 카페인 음료 마시기, 심한 신체 운동을 삼가 하도록 했다.

실험 결과, 의사결정 과제 수행 시 두 옵션(options) 중 한 옵션을 95% 이상 선택한 4명(두 조건에서 각각 2명)은 과제에서 요구되는 학습이 잘 이루어지지 않은 것으로 판단되어, 분석에서 제외시켰다. 따라서 이들을 제외한 총 52명(비교 통제 조건 26명, 스트레스 조건 26명)의 실험 결과를 분석했고, 비교 통제 조건과 스트레스 조건의 피험자들은 서로 연령 ($t_{50} = 1.23, p = .226$), 신체질량 지수(Body Mass Index; $t_{50} = -.20, p = .846$), 지난 한 달 간의 주관적 스트레스 정도($t_{50} = .71, p = .483$)에서 유의미한 차이가 없었다. 지난 한 달의 주관적 스트레스는 한국판 지각된 스트레스 척도(PSS-K, Korean Perceived Stress Scale; 백영묘, 2010). PSS-K는 최근 한 달의 주관적 스트레스를 측정하려는 목적으로 고안된 PSS(Cohen and Williamson, 1988)를 우리 말로 번안 및 표준화한 것이다.

모든 연구대상자들은 연구참여에 앞서 연구의 전반적인 절차에 대해

안내를 받았으며, 서면 동의서에 서명함으로써 연구참여에 대한 자발적인 의사를 밝혔다.

2) 스트레스 유발 실험

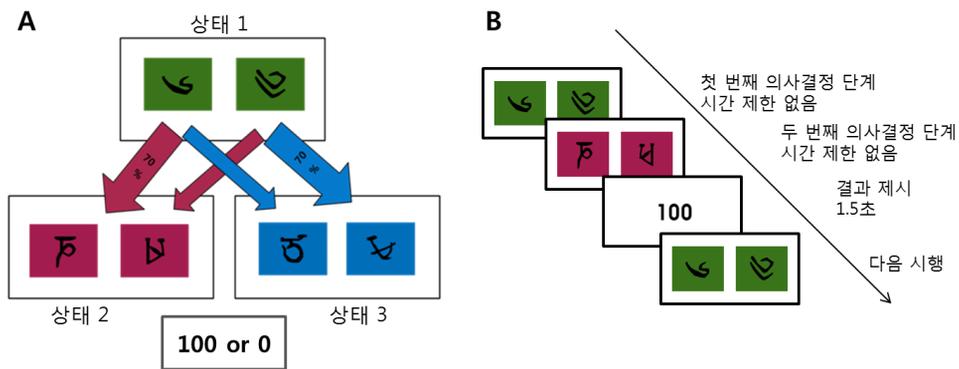
실험실에서 스트레스 실험 조건에 해당되는 연구참여자들에게 급성적으로 스트레스 반응을 유발시키기 위해서, 사회적 평가 상황에서의 냉 자극 압박 실험socially evaluated cold pressor test, SECPT (Schwabe et al., 2008; Schwabe and Wolf, 2009)을 실시했다. 실험 중 연구대상자들은 0 ~ 2°C의 차가운 얼음물에 한 손(오른손잡이는 왼손, 왼손잡이는 오른손)을 3분 이상, 혹은 버틸 수 있을 만큼의 시간 동안 (2명은 3분이 되기 전에 고통을 호소하여 2분 간만 유지했음) 손목까지 담그게 했으며, 그 과정은 낯선 사람을 통해 비디오 카메라로 녹화되었다. 비교 통제 조건의 연구참여자들은 36 ~ 38°C의 미지근한 물에 한 손을 손목까지 1분 간 담그게 했으며, 그 과정은 녹화하지 않았다. 두 조건의 처치가 각각 스트레스 반응 유발에 효과적이었는지 확인하기 위해, ‘전혀 스트레스를 느끼지 않았음’을 ‘0’, ‘매우 큰 스트레스’를 ‘100’으로 시각화한 시각적 아날로그 척도Visual Analogue Scale, VAS를 사용했다.

3) 의사결정 과제

두 단계로 이루어지는 반전 학습 과제_{2-stage reversal learning task} (그림 1-2)를 개발하여 사용했다. 이 과제는 선행 연구에서 개발한 두 단계 Markov 의사결정 과제_{2-stage Markov decision task} (Daw et al., 2011)를 변형하여 반전 학습 패러다임을 통합한 것으로, 의사결정자가 자신의 행동 선택과 그 결과 간의 관계가 반전되는 역동적이고 가변적인 상황에서 어떻게 행동을 선택해나가는지를 탐색하기 위해 개발되었다.

과제는 여섯 개의 구간_{block}으로 나뉘어 있고 각 구간은 40개의 시행들_{trials}로 이루어져, 총 240번의 시행으로 진행된다. 구간 간 쉬는 시간은 없으며, 구간이 바뀌는 것을 알아차릴 수 있는 단서는 주어지지 않는다. 의사결정자는 과제 중 화면을 보고 두 옵션_{option} (티베트 글자와 유사한 그림) 중 하나를 선택해야 하는데, 매번 점수를 받기 위해서는 총 두 단계_{stage}에 걸쳐 각각 한번씩의 선택을 해야 한다. 즉, 옵션_{option}의 색깔이 모두 초록색인 첫 번째 단계_{stage} (상태_{state} 1)에서 두 옵션_{option} 중 하나를 선택하면 옵션의 색깔이 모두 분홍색(상태 2)이거나 파란색(상태_{state} 3)인 두 번째 단계_{stage}로 바뀌며(즉 두 번째 단계_{stage}에 해당되는 ‘상태_{state}’는 두 종류임), 이 화면에서 두 번째 선택을 하고 나면 점수(0점 또는 100점)가 나타나는 방식이다. 즉 모든 시행_{trial}은 두 단계가 있으며(초록색 옵션_{option}들로 구성된 첫 번째 단계_{stage}와 분홍색 옵션_{option}들의 상태_{state} 2와 파란색 옵션_{option}들의 상태_{state} 3으로

구성된 두 번째 단계(stage), 첫 번째 단계에서 어떤 선택을 하는지에 따라 다음 단계가 상태_{state} 2가 될지 상태_{state} 3이 될지 결정된다(그림 1-2A).



C

	상태 2		상태 3	
	옵션 1	옵션 2	옵션 1	옵션 2
구간 1	0.6	0.2	0.6	0.2
구간 2	0.8	0.2	0.2	0.2
구간 3	0.2	0.2	0.8	0.2
구간 4	0.8	0.2	0.2	0.2
구간 5	0.2	0.2	0.8	0.2
구간 6	0.8	0.2	0.2	0.2

그림 1-2. 두 단계 반전학습 과제 (A) 과제의 구조(task structure). 첫 번째 단계에서의 선택이 다음 단계의 화면 상태와 확률적으로 연결되어 있고, 두 번째 단계에서의 각 옵션들은 각기 다른 보상확률에 따라 0점이나 100점의 결과로 연결된다. (B) 한 시행에서 일어나는 이벤트들의 타임라인. (C) 두 번째 단계에서의 옵션들에 대한 구간별 보상확률(reward probability)

첫 번째 단계_{stage}의 두 옵션_{option}들 중 하나는 분홍색 옵션들의 화면으로 전환될 가능성이 더 높고(상태_{state} 2로 전환될 확률 70%, 상태_{state} 3으로 전환될 확률 30%), 다른 하나는 그와 반대로 파란색 옵션들의 화면으로 전환될 가능성이 더 높다(상태_{state} 3으로 전환될 확률 70%, 상태_{state} 2로 전환될 확률 30%). 첫 번째 단계의 옵션들과 그 다음 단계의 ‘상태_{state}’ 간의 확률적 관계는 과제 중 변하지 않고 유지된다. 각 ‘상태_{state}’의 두 옵션들의 좌우 위치는 매 시행마다 무선적으로 바뀌게 했고, 두 번째 단계에서의 두 ‘상태_{state}’의 옵션 색깔(분홍 또는 파랑)은 연구참여자들마다 무선적으로 바꾸어서 색깔 효과를 상쇄시켰다.

두 번째 단계에서의 행동 선택은 정해진 확률에 따라 100점이나 0점의 점수로 연결되는데, 두 번째 단계_{stage}의 상태_{state} 2와 상태_{state} 3에서의 옵션들에 각각 주어진 보상확률_{reward probability}은 구간마다 정해진 규칙에 따라 바뀌게 된다(그림 1-2C). 구간_{block} 1에서는 상태_{state} 2의 두 옵션들이 100점과 연결될 확률이 각각 60%, 20%이며, 상태_{state} 3의 경우도 이와 같다. 따라서 구간_{block} 1에서는 두 상태_{state} (상태_{state} 2, 상태_{state} 3)의 기대치_{expected value}가 서로 같다. 그러나 이후 구간_{block} 2부터 마지막 구간_{block} (구간_{block} 6)까지는 두 번째 단계_{stage}의 옵션들의 보상 확률_{reward probability}이 반전 학습_{reversal learning}의 규칙을 따르게 된다. 즉, 구간_{block} 2에서는 상태_{state} 2의 옵션들의 보상 확률_{reward probability}이 각각 80%, 20%이나 상태_{state} 3의 옵션들은 모두 20%으로 설정되어, 두 상태_{state} 중 상태_{state} 2의 기대치가 높다. 그러나 구간_{block} 3에서는 80%의

보상확률이 상태 3의 옵션에 할당되어, 상태_{state} 2와 상태_{state} 3의 기대치에 반전이 일어나게 된다. 이러한 보상확률 설정으로, 구간_{block} 1에서는 첫 번째 단계_{stage}의 두 옵션 간 선호도의 차이가 발생하지 않지만, 구간_{block} 2 ~ 6에서는 상태_{state} 2와 상태_{state} 3 중 기대치_{expected value}가 더 큰 상태로 전환될 확률이 높은 옵션이 더 이롭다.

과제 수행 전, 연구참여자들에게 과제 첫 번째 단계_{stage}의 옵션들과 두 번째 단계_{stage}의 두 상태_{states} (상태 2, 상태 3) 간의 확률적 관계에 대해 설명했고, 두 번째 단계의 옵션들의 보상 확률은 과제 수행 중 지속적으로 바뀐다는 것을 설명했다. 과제 설명 후 과제의 진행에 익숙해질 수 있도록 연습 세션을 진행했다. 연습 과제는 연구참여자들이 과제의 구조를 파악하되 보상 확률 할당 방식을 미리 학습하지는 못하도록, 구간 1에서 6까지 각각 다섯 시행으로 구성하여 총 30시행을 진행했다.

4) 연구 절차

연구참여자로 선정된 피험자들은 선별 검사 후 1 ~ 2주 이내에 실험실을 방문하여 스트레스 유발 실험과 의사결정 과제 수행을 하도록 했다. 모든 실험은 스트레스 호르몬인 코티솔_{cortisol}의 일주기 리듬_{diurnal rhythm}이 안정화되는 오후 시간인 1시에서 5시 40분 사이에 진행했다.

피험자들은 실험실 입장 후 곧바로 스트레스 유발 실험에 대한 설명을 듣고 나서 스트레스 처치 혹은 통제 처치를 받았다. 처치가 끝나면 곧바로 실험자가 두 단계 반전 학습 의사결정 과제_{2-stage reversal learning decision task}에 대해 설명을 하고 나서 연습 세션을 진행했다. 연습 세션 후에는 본 과제를 실시하게 했으며, 본 과제 수행은 스트레스 처치 후 10분이 경과된 시점에 진행하도록 했다.

5) 행동 결과 분석

비교 통제 조건과 스트레스 조건 간의 스트레스 처치로 인한 주관적 스트레스 정도 및 과제 수행 차이를 검증하기 위해 양방향 t 검증(two-tailed t -test)을 실시했다. 두 조건의 수행을 대표하는 측정치로 t 검증에 사용한 측정치들은 다음과 같다. (1) 첫 번째 단계에서 두 옵션 중 하나를 고를 때까지 걸린 반응 시간의 평균, (2) 전체 수행 결과로 누적된 점수, (3) 첫 번째 단계에서 유리한 옵션(구간_{block} 2, 4, 6에서는 상태_{state} 2와 70%로 연결된 옵션; 구간_{block} 3, 5에서는 상태_{state} 3과 70%로 연결된 옵션)을 선택한 확률. 다음으로는 피험자들의 (첫 번째 단계에서의) 이전 행동 유지 확률에 대한 요인을 세 가지로 설정하여, 3요인 혼합 설계 분산 분석_{mixed-design ANOVA}을 실시했다. 피험자 내 요인 두 가지로 (1) 직전 수행의 보상 여부와 (2) 단계 전환 형태_{stage transition}

type(첫 번째 단계의 옵션 선택의 결과, 높은 확률 관계인 ‘상태_{state}’로 전환되었는지, 낮은 확률 관계인 ‘상태_{state}’로 전환되었는지)를 설정했고, 피험자 간 요인으로는 스트레스 처치 여부를 설정했다. 마지막으로 반전 학습_{reversal learning} 구간_{block}의 행동-결과 수반성_{contingency} 변화에 대한 유연한 대처에 스트레스 효과가 유의한지를 검증하기 위해, 구간 3 ~ 6에서 첫 번째 단계의 유리한 옵션 선택 확률에 대해 피험자 내 요인으로 구간 내 하위 구간_{sub-block}(전반부_{initial half}: 시행_{trial} 1 ~ 20; 후반부_{final half}: 시행_{trial} 21 ~ 40), 피험자 간 요인으로 스트레스 처치 여부를 설정하여 혼합 설계 분산분석을 실시했다.

6) 의사결정 행동의 계산 모형화

행동 결과 분석에 이어, 각 피험자 마다 과제 수행으로 축적된 매 시행의 행동-결과 자료에 대해 계산적 강화학습_{reinforcement learning} 모형을 적용했다. 세 가지 강화학습 모형을 사용했고, 각각은 다음과 같다. (1) 모형 기반_{model-based} 선택 행동에 대한 가치 함수와 모형 부재_{model-free} 선택 행동에 대한 가치 함수_{value function} 간의 가중치 결합 모형 (RL model 1), (2) 모형 기반_{model-based} 행동 경향과 모형 부재_{model-free} 행동 경향이 각각 두 모수_{parameter}로 반영되고 행동 가치가 하나의 감소 모수_{decay parameter}에 의해 업데이트되는 모형(선택된 행동과 선택되지

않은 행동이 모두 업데이트 됨) (RL model 2), (3) 모형 기반_{model-based} 행동과 모형 부재_{model-free} 행동이 두 종류의 모수로 구분되며, 두 종류의 감소 모수_{decay paramter}로 업데이트되는 모형 (RL model 3).

RL model 1. 이 모형은 선행 연구(Daw et al, 2011)에서 고안된 것으로, 의사결정자가 매번 행동을 선택할 때마다 모형 기반_{model-based} 가치 계산과 모형 부재_{model-free} 가치 계산이 서로 선형적으로 트레이드 오프_{trade-off}를 이루어 작용한다는 가정을 한다. 모형 기반의 가치 계산_{model-based valuation, MBV}이 벨만 방정식_{Bellman's equation}으로 구현되고, 내적 모형 없이 이루어지는 가치 계산_{model-free valuation, MFV}은 시간차 학습_{temporal difference learning model}로 구현되어, 매 회의 의사결정 행동이 MBV과 의 MFV 간 트레이드 오프를 통해 업데이트된다.

구체적으로는, 과제의 첫 번째 단계(S1)에서의 의사결정 행동에 대한 MFV와 MBV를 각각 V_{S1}^{MF} , V_{S1}^{MB} 로 나타낸다면, t 번째 시행에서 첫 번째 단계의 두 옵션에 대한 기대 값은 각각 $V_{S1,t}(1)$, $V_{S1,t}(2)$ 로 표현할 수 있으며, 그 다음 시행 ($t+1$)에서의 기대 값은 다음과 같은 형태의 수학적 알고리즘을 통해 구현된다.

$$V_{S1,t+1}(\mathbf{x})^{\text{Hybrid}} = \omega V_{S1,t}(\mathbf{x})^{\text{MB}} + (1-\omega) V_{S1,t}(\mathbf{x})^{\text{MF}}, \quad (5)$$

이 식에서 ω 는 MBV 식과 MFV 식을 선형 조합하는 가중치이다.

과제의 두 번째 단계에서는 상황에 대한 심적 모형 없이 보상 예측 오류 reward prediction errors, RPE 에 의해 행동 선택이 이루어질 것이 예상되므로, 상태 2 를 S2, 상태 3 을 S3 이라 할 때 각 화면마다의 두 옵션들의 기대값 expected value 은 다음과 같은 식을 통해 업데이트된다.

$$\begin{aligned} V_{S2,t+1}(\mathbf{x}) &= V_{S2,t}(\mathbf{x}) + \alpha_2(R - V_{S2,t}(\mathbf{x})) \text{ 또는} \\ &= V_{S3,t}(\mathbf{x}) + \alpha_2(R - V_{S3,t}(\mathbf{x})), \end{aligned} \quad (6)$$

이 식에서 α_2 는 두 번째 단계에서 매 회 RPE 가 가치함수 업데이트에 반영되는 정도를 결정하는 학습률 learning rate 이며, R 은 의사결정 행동의 실제 결과로 보상을 받거나 못 받음을 반영하는 관찰치이다.

첫 번째 단계에서 선택한 옵션의 MFV 는 시간차 학습 방식에 따라 업데이트되므로, 첫 번째 단계의 의사결정으로 인한 다음 단계에서의 RPE 의 영향은 모수 λ eligibility trace (Sutton & Barto, 1998)로 인해 감소되어 반영된다(식 7).

$$\begin{aligned} V_{S1,t+1}^{\text{MF}}(\mathbf{x}_{S1}) &= V_{S1,t}^{\text{MF}}(\mathbf{x}_{S1}) + \alpha_1(V_{S2,t}(\mathbf{x}_{S2}) - V_{S1,t}^{\text{MF}}(\mathbf{x}_{S1})) + \lambda \alpha_1(R - V_{S2,t}^{\text{MF}}(\mathbf{x}_{S2})) \text{ 또는} \\ V_{S1,t+1}^{\text{MF}}(\mathbf{x}_{S1}) &= V_{S1,t}^{\text{MF}}(\mathbf{x}_{S1}) + \alpha_1(V_{S3,t}(\mathbf{x}_{S3}) - V_{S1,t}^{\text{MF}}(\mathbf{x}_{S1})) + \lambda \alpha_1(R - V_{S3,t}^{\text{MF}}(\mathbf{x}_{S3})), \end{aligned} \quad (7)$$

이 식에서 $\mathbf{x}_{S1} \in \{a_1, a_2\}$, $\mathbf{x}_{S2} \in \{a_3, a_4\}$, $\mathbf{x}_{S3} \in \{a_5, a_6\}$ 이며, α_1 와 α_2 는 각각 첫 번째 단계와 두 번째 단계의 학습률 learning rate 이다.

첫 번째 단계에서 선택한 옵션에 대한 MBV는 본 과제의 단계 전환

확률을 반영한 벨만 방정식(Bellman's equation)을 사용하여 다음과 같은 식으로 구현된다.

$$V_{S_{1,t+1}}(a_1)^{MB} = P(S_2 | S_{1,t}, a_1) \max(V_{2,t}(a_3), V_{2,t}(a_4)) + P(S_3 | S_{1,t}, a_1) \max(V_{3,t}(a_5), V_{3,t}(a_6)),$$

$$V_{S_{1,t+1}}(a_2)^{MB} = P(S_2 | S_{1,t}, a_2) \max(V_{2,t}(a_3), V_{2,t}(a_4)) + P(S_3 | S_{1,t}, a_2) \max(V_{3,t}(a_5), V_{3,t}(a_6)) \quad (8)$$

첫 번째 단계에서 의사결정자가 옵션 a_1 을 선택할 확률 $P(a_1)$ 는 두 옵션 간의 기대치 차이와 이전 선택의 보속성을 함께 반영할 수 있도록 소프트맥스 함수(softmax function)로 계산했다. 식 (9)에서 π 는 보속성을 반영하는 모수이며(C 는 이전 시행에서 현 시행과 같은 선택을 했는지 여부를 반영, 같은 선택이었으면 $C=1$, 아니면 $C=0$), β 는 의사결정자의 무선적 행동 탐색 경향을 반영하는 모수(inverse temperature)이다.

$$P_{S_{1,t+1}}(a_1) = 1 / (1 + \exp(-\beta ((V_{S_{1,t+1}}(a_1) - V_{S_{1,t+1}}(a_2)) + \pi (C_{S_{1,t}}(a_1) - C_{S_{1,t}}(a_2))))), \quad (9)$$

즉, RL model 2에서는 α_1 , α_2 , β , λ , ρ , ω 총 여섯 개의 모수가 추정되며, 과제 중 매 시행의 첫 번째 단계 의사결정에 대한 우도(likelihood)만으로 최대우도(maximum likelihood)가 계산된다.

RL model 2. RL model 2는 RL model 1과 달리 MFV와 MBV가 서로 독립적으로 이루어진다는 가정을 한다. 이 모형은 선행 연구에서 사용된 Rescolar-Wagner 강화학습 모형의 변형된 알고리즘(Barraclough et

al., 2004) (식 10)을 다시 수정한 것이다. 두 개의 옵션이 있는 의사결정 상황에서 행위자가 두 옵션 중 하나를 선택할 때, t 번째 시행에서 선택한 옵션 x 에 대한 가치 함수를 $V_t(x)$ 로 표현하면, $V_{t+1}(x)$ 은 Rescolar-Wagner 식을 변형하여 식 (10)과 같이 나타낼 수 있다.

$$\begin{aligned} V_{t+1}(x) &= (1-\alpha)V_t(x) + \alpha \kappa_+, && \text{보상 받은 경우,} \\ &= (1-\alpha)V_t(x) + \alpha \kappa_-, && \text{보상 받지 못한 경우} \end{aligned} \quad (10)$$

이때 α 는 선택한 옵션의 기대치 변화에 가중되는 학습률(learning rate)이자, 선택하지 않은 옵션에 대한 망각률(discount factor)의 역할을 한다. 이 모형에서 모수 κ 는 선택된 옵션의 가치의 변화량으로, κ_+ 는 보상을 받았을 때의 강화 정도이고, κ_- 는 보상을 받지 못했을 때의 부정적인 영향의 정도를 반영한다. 이러한 강화학습 모형은 선행 연구에서 동물의 선택 행동을 잘 반영하는 것으로 입증된 바 있다(Ito and Doya, 2009).

본 연구에서는 식 (10)을 확장하여 의사결정 중 이루어지는 독립적인 MBV, MFV 과정을 모수 κ 에 반영하였고, κ 를 모형 기반 변화량 κ^{mb} 와 모형 부재 변화량 κ^{mf} 으로 확장시켰다. 즉 본 연구에서 사용된 RL model 2에서는 첫 번째 단계의 옵션 선택에 대한 가치 함수 업데이트가 단계 간 전환 형태에 따라 달라진다. 예를 들어 70%의 확률로 전환된 '상태'에서 보상을 받으면, 선택한 옵션의 가치함수는 $\kappa_+^{mf} + \kappa_+^{mb}$ 에 의해 업데이트된다. 이 경우에는 선택 행동의 긍정적

결과가 MFV와 MBV 모두에 기인한다고 가정할 수 있기 때문이다. 이에 반해, 30%의 가능성으로 전환된 '상태'에서 보상을 받으면, 선택한 옵션의 가치함수는 κ_{+}^{mf} 에 의해서 업데이트되고, 선택하지 않은 옵션의 가치함수는 κ_{+}^{mb} 에 의해서 업데이트 된다. 30%로 전환된 '상태'에서의 보상은 선택한 옵션의 가치를 MFV 과정을 통해 변화시키고, 선택하지 않은 옵션의 부정적인 가치를 MBV 과정을 통해 변화시킬 것이 예상되기 때문이다. 이와 같은 방식으로, 70%의 확률로 전환된 '상태'에서 보상을 받지 못했으면, 선택한 옵션은 $\kappa_{-}^{mf} + \kappa_{-}^{mb}$ 에 의해 업데이트 되고, 30%의 확률로 전환된 '상태'에서 보상을 받지 못했으면 선택한 옵션은 κ_{-}^{mf} 에 의해, 선택하지 않은 옵션은 κ_{-}^{mb} 에 의해 업데이트 된다.

이 모형을 실제 실험 자료에 적용할 때에는 아래의 식 (11)이 적용되었다. 식 (10)을 사용한 모형 추정 과정에서는 알고리즘의 반복 연산_{iteration} 결과, 학습률이 0에 매우 가깝게 추정되는 피험자의 자료에 대해서는 κ 값이 매우 크게 추정될 것이 예상되며, 실제 모형 추정 결과, 학습률이 매우 작게 나타나는 피험자들에게서 κ 값이 매우 크고 불안정적인 값으로 추정되었다. 따라서 식 (10)과 수학적으로 동일한 식 (11)이 사용되었다.

$$\begin{aligned}
 V_{t+1}(\mathbf{x}) &= \gamma V_t(\mathbf{x}) + \mathcal{L}_+, && \text{보상 받은 경우,} \\
 &= \gamma V_t(\mathbf{x}) + \mathcal{L}_- && \text{보상 받지 못한 경우,} \quad (11)
 \end{aligned}$$

이 식에서 γ 은 학습률을 1에서 뺀 값($\gamma = 1 - \alpha$)이며, 선택한 옵션에 대한 이전 기대치에 부과되는 모수 parameter 로, 이전 경험이 다음 행동 선택에 미치는 영향의 정도를 나타낸다. Δ 는 α 와 κ 의 곱($\Delta = \alpha \kappa$)으로, 의사결정자의 행동 선택과 그 결과에 따라 결정되는 가치함수의 변화량을 반영한다. 따라서 RL model 2에서 Δ_+^{mf} , Δ_+^{mb} , Δ_-^{mf} , Δ_-^{mb} 는 각각 $\alpha \kappa_+^{mf}$, $\alpha \kappa_+^{mb}$, $\alpha \kappa_-^{mf}$, $\alpha \kappa_-^{mb}$ 와 수학적으로 동일하다.

RL model 2에서는 Δ_- 값이 0보다 작을 때에는 보상 받지 못한 행동을 다음 시행에서 다른 행동 선택으로 바꾸려는 경향을 반영하고, 0보다 클 때에는 보상 받지 못했더라도 이전 행동을 다음에도 유지하는 경향을 반영한다. 즉 Δ_-^{mf} 와 Δ_-^{mb} 는 부적인 값을 가질 때에는 그 값이 클수록 각각 보상 받지 못했을 때 다음 시행에서 행동을 수정하려는 모형 부재 행동 경향과 모형 기반 행동 경향을 나타낸다. 따라서 RL model 2에서는 두 델타 값, Δ_+ 과 Δ_- 간의 차이($\Delta_+ - \Delta_-$)가 의사결정자의 행동 선택 시 MFV와 MBV의 기여도를 의미하게 된다. 즉 $\Delta_+^{mf} - \Delta_-^{mf}$ 와 $\Delta_+^{mb} - \Delta_-^{mb}$ 는 각각 MFV, MBV의 정도를 반영한다.

첫 번째 단계에서 각 옵션을 선택할 확률은 두 옵션의 기대치 차이에 기반하여 소프트맥스 함수 softmax function에 의해 계산했다.

$$P_t(a_1) = 1 / (1 + \exp(-(V_t(a_1) - V_t(a_2)))) \quad (12)$$

식 (12)에서 a_1, a_2 는 각각 옵션 1 을 선택하는 행동과 옵션 2 를 선택하는 행동을 나타낸다.

한편 RL model 2 는 델타 값들에 의해 각 옵션의 기대치가 조절되므로, 의사결정자가 무선적으로 행동을 탐색하는 경향성을 반영하는 모수 parameter 는 별도로 설정하지 않았다.

RL Model 3. 마지막 강화학습 모형은 RL model 2 와 같으나, 이전 경험의 영향의 정도를 반영하는 모수 parameter 가 두 종류(γ_{common} and γ_{rare})로 구분되는 모형이다. 이 모형은 30%의 확률로 전환된 상태에서의 학습률($1 - \gamma_{\text{rare}}$)은 70%의 확률로 전환된 상태에서의 학습률($1 - \gamma_{\text{common}}$)과 구분되며 후자보다 전자에 모형 기반의 속성이 크다는 가정을 한다.

모형 적용. Matlab (Mathworks, Inc) 기반의 우도 likelihood 추정 함수 fminsearchbnd 를 이용하여 피험자 개인의 과제 수행 자료들에서 로그화된 우도 likelihood 값을 최대화하는 방식으로 모형의 모수치들을 추정했다. RL model 1 에서는 학습률 α_1, α_2 의 추정 범위를 0 에서 1 사이로 제한했으며, 그 외에 다른 모수들의 추정 범위는 제한하지 않았다. RL model 2 와 RL model 3 에서는 γ 만 0 에서 1 사이의 범위 내에서 추정하고 다른 모수들은 추정 범위를 제한하지 않았다. 모수 추정 parameter estimation 시에는 모수마다 무선적으로 선택된 초기값으로

사용하여 우도 함수 likelihood function 내에서 지역적인 최대값을 추정할 가능성 local maxima problem 를 피하고 함수 전체의 최대값을 구할 수 있을 때까지 추정 과정을 반복했다.

모형 비교. 상기 세 모형 추정 결과를 베이저안 정보 기준 Bayesian information criteria, BIC (Schwarz, 1978)와 아카이케 정보 기준 Akaike information criteria, AIC (Adcock et al., 2006)를 사용하여 비교했다. BIC와 AIC는 모두 모형 선택 시 사용되는 모형의 적합도 지표로, 각 모형의 우도 함수 likelihood function 에 기반하여 계산된다. 행동 자료에 모형을 적용시킬 때 모수 parameter 개수가 증가하면 우도 likelihood 가 개선되기 마련이므로, 모형이 행동 자료에 적절하지 않을 때에도 모수 parameter 가 많으면 우도 likelihood 가 개선되는 경우가 있다. 따라서 BIC와 AIC는 모두 모형의 모수 parameter 개수에 따라 지표 값이 증가하도록 계산되어 모형에 패널티 penalty 를 부여하며, BIC의 경우 AIC에 비해 모수 parameter 개수로 인한 패널티가 크다.

표 1-1. 모형 비교 지표 값의 평균 (표준오차)

	부적 로그 우도 negative log likelihood	BIC	AIC
RL model 1	114.26 (5.93)	261.40 (11.87)	240.52 (11.87)
RL model 2	114.43 (5.94)	256.27 (11.89)	238.87 (11.89)
RL model 3	114.02 (5.91)	260.92 (11.83)	240.04 (11.83)

BIC: Bayesian information criterion; AIC: Akaike information criterion

모형 비교 결과, RL model 1 은 다른 두 모형에 비해 행동 자료에 적합하지 않은 것으로 나타났다(표 1-1). 즉, 대응표본 t 검증 결과, RL model 1 의 AIC 평균 값은 RL model 2 에 비해 높았지만 그 차이가 통계적으로 유의하지는 않았다($t = 1.28, p = .206$). 그러나 BIC 값은 RL model 1 이 RL model 2 에 비해 유의한 수준으로 높게 나타났으며($t = 3.97, p < .001$), 이는 RL model 1 이 모수의 개수가 더 많음에도 불구하고 RL model 2 에 비해 행동에 대한 적합도가 개선되지 않았음을 뜻한다. 또한 RL model 1 은 행동 자료에 적용한 결과 β 가 매우 크게 추정되거나 ω 값이 부적인 값으로 추정되는 등의 문제가 나타났으며, 이는 모형이 행동 결과에 적합하지 않음을 뜻한다(Daw, 2011). 따라서 RL model 1 이 본 연구의 피험자들의 의사결정 과제 수행 결과에 맞지 않는 것으로 보여, 모수 추정 parameter estimation 결과를 분석에서 제외시켰다.

한편 RL model 2 는 RL model 3 에 비해서도 적합도가 좋은 것으로 나타나(BIC: $t_{51} = 19.24, p < .001$; AIC: $t_{51} = 4.84, p < .001$), RL model 2 이 피험자들의 의사결정 과제 수행을 가장 잘 반영하는 것으로 판단된다.

2. 연구결과

1) 행동 분석 결과

스트레스 처치 조건과 통제 조건 각각 26 명의 총 52 명에 대한 두 단계 반전 학습 의사결정 과제 수행 자료를 분석했다(표 1-2).

표 1-2. 실험 결과의 조건 별 평균(표준 오차)과 조건 간 차이

	주관적 스트레스(VAS*)	반응시간	과제 수행 총점	유리한 옵션 선택 확률
통제조건	0.56(0.15)	1.13(0.11)	11588.46(376.31)	0.61(0.02)
스트레스조건	6.74(0.70)	0.91(0.08)	10357.69(311.48)	0.53(0.02)
T 값 (p)	-8.61(<.001)	1.60(.117)	2.52(.015)	2.90(.006)

* visual analogue scale

예상했던 바와 같이, 스트레스 조건의 피험자들은 통제 조건 피험자들에 비해 처치 (통제 처치 또는 스트레스 처치)로 인한 주관적 스트레스를 높게 보고했고(two-tailed t -test: $t_{50} = 8.61$, $p < 0.001$), 두 조건 모두 성별 차이는 유의하지 않았다($p > 0.12$). 또한 첫 번째 단계에서의 의사결정의 평균 반응 시간에서도 조건 간 차이는 유의하지 않았다 (two-tailed t -test, $t_{50} = 1.6$, $p = .117$). 그러나 스트레스 조건의 피험자들은 과제 수행으로 얻은 점수의 총합이 통제 조건보다 낮았으며($t_{50} = 2.52$, $p = .015$), 구간마다 옵션들의 보상 확률이 바뀌는 구간 3 ~ 6 에서 유리한 옵션을 선택한 확률 역시 통제 조건에 비해

유의하게 저조했다($t_{50} = 2.90, p = .006$). 이러한 결과는 스트레스가 보상을 최대화하는 목표지향적 행동에 부정적인 영향을 미쳤을 가능성을 시사한다.

반전 학습 양상 분석. 본 과제는 구간_{block} 2에서부터 구간_{block} 6까지 매 구간마다 예고나 단서 없이 첫 번째 단계의 두 옵션 간 가치의 우월성이 반전되므로, 피험자들은 해당 구간들에서 행동-결과 수반성의 급작스러운 반전 reversal 을 경험하기 시작한다. 행동과 결과 간의 수반성 contingency 변화는 의사결정의 상황 변화에 해당하므로, 모형 기반의 행동 선택을 하는 의사결정자라면 이러한 변화에 대한 정보를 받아들여 행동 선택에 반영할 것이며, 그 결과로 구간_{block} 3 ~ 6에서는 구간 내의 전반부에서는 행동-결과 수반성 contingency 변화로 인해 유리한 행동 선택 확률이 떨어졌다가 후반부에는 높아질 것이 예상된다. 스트레스가 행동 선택 시 목표지향 체계의 영향력을 저하시킨다면, 행동-결과 수반성 contingency 반전 reversal 이 반복되는 구간_{block} 3 ~ 6에서 변화에 유연하게 대처하는 것을 방해할 것이 예상된다. 따라서 이번에는 구간_{block} 3 ~ 6의 각 구간_{block} 내 전반부 시행들(시행 1 ~ 20)과 후반부 시행들(시행 21 ~ 40)에서 첫 번째 단계의 유리한 옵션(이를 테면 구간 3, 5에서는 옵션 2, 구간 4, 6에서는 옵션 1)을 선택하는 확률을 각각 구분하여 계산했다. 즉 첫 번째 단계에서 두 옵션 중 가치가 우월한 것을 선택하는 확률에 피험자 내 요인으로 구간 내의 하위 구간(전반부,

후반부)을 지정하고, 피험자 간 요인으로 스트레스 처치 유무를 지정하여 혼합 설계 분산분석을 실시하였다.

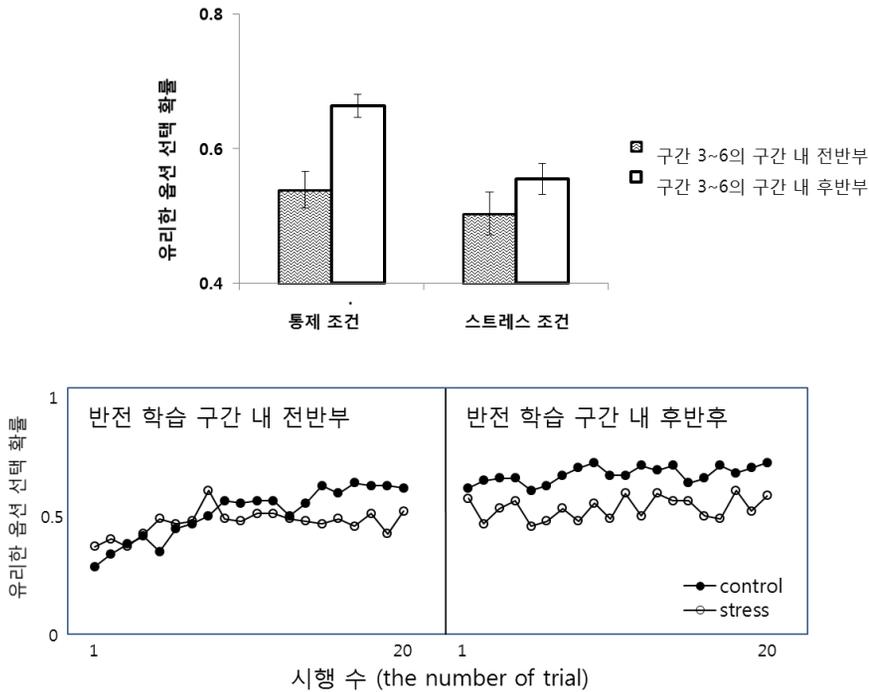


그림 1-3. 반전학습(reversal learning)에 대한 스트레스의 영향 (A) 구간 3 ~ 6에서 유리한 옵션을 선택하는 행동 비율에 대한 구간 내의 하위 구간 종류(전반부, 후반부)와 스트레스 처치 유무 간의 상호 작용 효과가 유의하게 나타남. 오류 막대는 평균의 표준 오차. (B) 구간 3 ~ 6 내에서 시행 수의 진행에 따라 유리한 옵션을 선택하는 비율의 변화양상. 네 개의 구간에서의 수행을 조건 별로 평균 냄.

분석 결과, 구간 내 하위 구간의 주효과가 유의한 것으로 나타나($F_{(1,50)} = 28.53, p < .001$) 본 과제 수행에 반전 학습의 특성이

잘 반영된 것으로 보이며, 스트레스의 주효과 역시 유의하여($F_{(1,50)} = 5.00, p = .030$), 과제 전반적으로 스트레스가 유리한 의사결정을 방해한 것으로 보인다(그림 1-3). 특히 예상했던 바와 같이, 분석 결과에서 두 요인 간의 상호작용 효과가 유의하게 나타났으며($F_{(1,50)} = 4.98, p = .030$), 이는 스트레스가 반전 학습 reversal learning 을 방해했음을 시사한다.

이전 행동 유지-전환 양상 분석. 스트레스가 의사결정 행동 과정 중 습관 체계와 목표지향 체계의 기여도를 각각 어떻게 변화시켰는지 탐색하기 위해, 첫 번째 단계의 의사결정 시 매 시행마다의 행동 유지 혹은 전환 양상을 분석했다. 모형 부재 model-free 행동 선택 체계, 즉 습관 체계에 의해 행동을 선택하는 의사결정자라면, 매 시행에서 이전 행동을 유지할지 전환할지 여부가 이전 시행의 행동이 보상을 받았는지 못 받았는지에 따라 정해될 것이다. 반면 모형 기반 model-based 행동 체계, 즉 목표지향적 체계를 기반으로 행동을 선택하는 의사결정자라면 과제의 전체적인 구조에 대한 심적 표상을 바탕으로 최적의 행동을 선택할 것이므로, 매 시행에서 의사결정을 할 때 이전 시행에서 자신이 선택한 옵션이 어떤 단계 전환 형태를 거쳤는지(70%의 확률 관계인 화면으로 전환되었는지, 30%의 관계인 화면으로 전환되었는지)를 고려할 것이다. 즉, 70%의 관계인 화면으로 전환되어 100 점을 받았다면 그 옵션을 선택한 행동은 다음 시행에서도 유지할 것이고, 30%의 관계인 화면으로

전환되어 100 점을 받았다면 다음 시행에서는 다른 옵션을 선택하게 될 것이다. 이는 심적 모형이 부재한 행위자가 단계의 전환 확률에 대한 고려 없이 보상받은 행동을 유지하고 그렇지 못한 행동은 다른 행동으로 전환할 것으로 예상되는 것과 대조적이다(그림 1-4A).

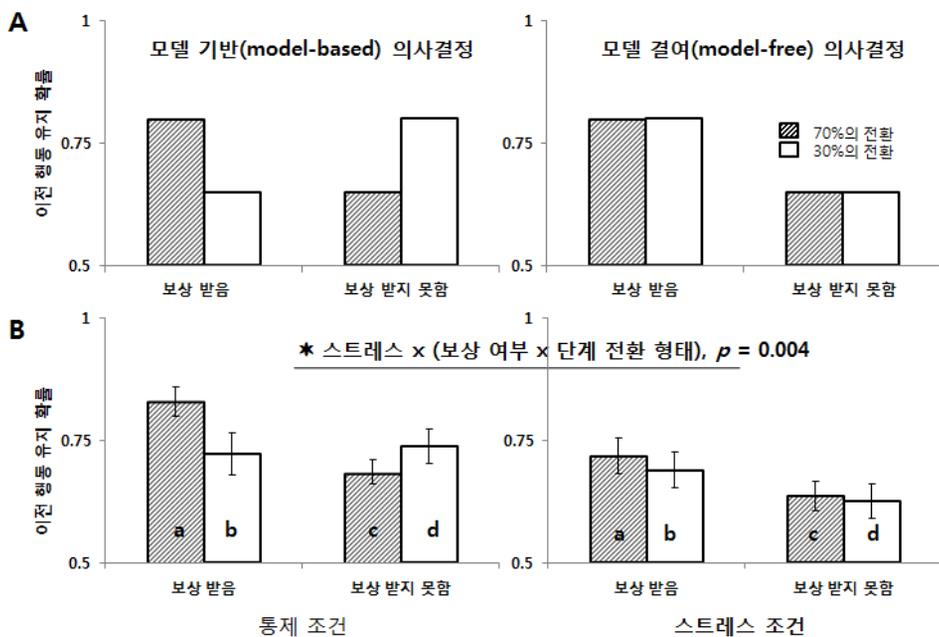


그림 1-4 행동 선택 유지-전환 양상 분석 (A) 모형 기반model-based 의사결정자의 행동 유지 양상에 대한 가설적 도식(왼쪽)과 모형 부재model-free 의사결정자의 행동 유지 양상에 대한 가설적 도식(오른쪽) (B) 통제 조건(왼쪽)과 스트레스 조건(오른쪽)의 피험자들의 실제 과제 수행 양상. 스트레스 x 행동의 결과 x 과제 구조 간의 상호작용 효과가 유의하게 나타남($p < .005$). 오류 막대는 평균의 표준오차를 의미함.

이러한 예측 하에서는 의사결정자가 과제 첫 번째 단계에서 이전 행동을 다음에도 유지할 확률에 두 가지 피험자 내 요인, (1) 이전

행동의 결과인 보상 여부와 (2) 과제 구조, 즉 이전 행동으로 인한 단계 전환 형태(70%의 전환 또는 30%의 전환)를 가정할 수 있다. 또한 피험자 간 요인으로서 스트레스 처치 여부를 가정하여, 스트레스가 결과 기반의 습관적 행동 선택과 과제 구조에 대한 지식을 이용하는 목표지향적 행동 선택에 미치는 영향을 탐색하고자 했다.

혼합 설계 분산분석 mixed design ANOVA 결과(그림 1-4B), 이전 행동의 결과(보상 여부)의 주효과가 유의한 것으로 나타났고($F_{(1,50)} = 18.44, p < 0.001$), 이는 과제 수행에 모형 없이 model-free 행동을 선택하는 습관 체계가 영향을 미쳤음을 시사한다. 또한 행동의 결과(보상 여부)와 과제 구조 task structure 간의 상호작용이 유의미하게 나타났고($F_{(1,50)} = 14.06, p < 0.001$) 이는 과제 수행에 모형 기반 model-based 행동 체계가 영향력을 행사했음을 시사한다. 이러한 두 요인 간 상호작용 효과와 스트레스 간의 상호작용이 역시 유의하게 나타나, 스트레스가 모형 부재 model-free 행동 체계와 모형 기반 model-based 행동 체계의 영향력에 조절자 modulator 의 역할을 한 것으로 보인다($F_{(1,50)} = 8.86, p = 0.004$). 한편 스트레스와 보상 여부 간의 상호작용($F_{(1,50)} = 0.06, p = 0.81$)과 스트레스와 단계 전환 형태 간의 상호작용($F_{(1,50)} = 0.09, p = 0.76$)은 유의하지 않았다.

스트레스, 행동의 결과(보상 여부), 과제 구조(단계 전환 종류) 간의 상호 작용에 대한 특성을 보다 면밀히 탐색하기 위해, 두 조건의 모형 기반 수행의 지표(model-based index, $MBI = (a + d) - (b + c)$), 그림 1-4B)와 모형 부재 수행의 지표(model-free index, $MFI = (a +$

b) - (c + d), 그림 1-4B)를 계산하여, 조건 간 차이를 검증했다. MFI 는 첫 번째 단계에서의 의사결정 시 직전 행동이 보상을 받았으면 같은 행동을 유지하고 그렇지 않으면 유지하지 않는 행동 양상을 수치화하는 지수인 반면, MBI 는 첫 번째 단계에서의 의사결정 시 직전 행동의 보상 여부가 두 번째 단계에서 어떤 화면을 통한 결과인지를 고려하여 행동을 유지하거나 바꾸는 양상을 반영하는 지수이다. 양방향 *t*-검증 결과, 스트레스 조건의 피험자들은 통제 조건의 피험자들에 비해 MBI 값이 낮게 나타났으며($t_{50} = 2.98, p = .004$) 이는 스트레스가 모형 기반 model-based 의 의사결정 과정을 방해했음을 뜻한다. 한편 모형 부재 model-free 행동 경향을 반영하는 행동 지표 MFI 값에서는 두 조건 간 유의미한 차이가 나타나지 않았다($t_{50} = -.24, p = .812$).

2) 모형 적용 결과

행동 분석 결과는 스트레스가 의사결정 시 모형 기반 model-based 행동 과정의 기여도를 감소시켰음을 시사한다. 그러나 행동 유지-전환 양상 분석이나 총점 및 유리한 행동 선택 확률에 대한 조건 간 차이 분석은, 직전 시행의 행동과 그 결과에 의해 일어나는 행동 선택 양상을 탐색하거나 각 조건 간 수행의 평균치에 근거한 특성을 탐색할 수 있을 뿐, 과제 수행 전체에서 매 시행마다 행동-결과의 경험이 누적되어

행동에 미치는 영향을 반영하지는 못한다. 강화학습 reinforcement learning 의 이론적 가설에서는 현재 시행에서의 행동 선택에 행위자 agent 가 이전에 경험한 모든 행동-결과 관계가 반영되며, 현재의 행동 역시 이후의 행동 선택에 지속적으로 영향력을 갖는다. 또한 강화학습 reinforcement learning 에 대한 계산적 접근 시, 계산 모형 알고리즘에 의사결정 행동의 가설적 요인을 반영하여 행위자의 개인적 행동 특성을 양적으로 추정할 수 있다. 따라서 본 연구에서는 행동 분석 다음으로 피험자들의 수행 자료, 즉 과제 내 240 번의 시행에서의 행동 선택과 그 결과에 대해 축적된 자료에 가설적 강화학습 모형 reinforcement learning model 을 적용하여, 개인적 행동 특성을 추정했고 그에 대한 스트레스의 영향을 검증했다.

RL model 2. RL model 2에서는 모형 부재 model-free 처리과정과 모형 기반 model-based 처리과정 간의 관계가 상보적인 관계로 가정되지 않고 각각 독립적인 모수 parameter 로 모형 알고리즘에 반영된다. 이 모형에서는 선택 행동의 기대치 업데이트가, 그 행동 결과로 보상을 받는지 받지 못하는지에 따라 각기 다른 이론적 모수 parameter 에 의해 이루어진다. 따라서 모형 기반 model-based 의 행동 체계의 영향과 모형 부재 model-free 행동 체계의 영향은 각각 보상 받을 때의 변화 모수 parameter 와 보상 받지 못할 때의 변화 모수 parameter 간의 차이, 즉 $\Delta_+ - \Delta_-$ 로 간주된다. 이 모형의 모수 최대우도추정치 maximum likelihood estimate 를 구한 결과(표 1-3), 스트레스 조건의 피험자들은 통제 조건에 비해 Δ_+^{mb} 와 Δ_-^{mb} 간 차이,

즉 $\Delta_+^{mb} - \Delta_-^{mb}$ 값이 유의하게 낮았고(양방향 t -검증, $t_{50} = 2.12$, $p = 0.039$), Δ_-^{mf} 값은 좀 더 부적인 값으로 추정되었다($t_{50} = 3.52$, $p = 0.001$). 이는 스트레스가 모형 기반 $_{model-based}$ 행동을 방해했을 가능성을 시사하며, 동시에 보상을 받지 못한 행동에 대해서는 모형 부재 $_{model-free}$ 방식으로 행동 전환하는 양상을 증진시켰음을 시사한다. 피험자들마다 추정된 Δ_-^{mf} 값들은 $\Delta_+^{mb} - \Delta_-^{mb}$ 와 상관관계가 유의미하지 않은 것으로 나타나($r = .23$, $p = .095$), 스트레스가 모형 기반 $_{model-based}$ 의 행동 과정에 미치는 영향과 모형 부재 $_{model-free}$ 행동 과정에 미치는 영향은 서로 독립적인 것으로 보인다.

표 1-3. 강화학습 모형에서의 조건 별 모수 추정치 사분위 값들

	RL model 2					RL model 3					
	γ^*	Δ_+^{mf}	$\Delta_+^{mb\dagger}$	Δ_-^{mf*}	$\Delta_-^{mb\dagger}$	γ_{common}	γ_{rare}^*	Δ_+^{mf}	Δ_+^{mb*}	Δ_-^{mf*}	Δ_-^{mb*}
통제 조건											
25 th	0.23	0.70	0.10	0.12	-0.41	0.28	0.03	0.62	0.10	0.10	-0.64
Median	0.47	1.09	0.33	0.48	-0.14	0.50	0.32	1.03	0.25	0.44	-0.26
75 th	0.60	1.67	0.80	0.78	0.04	0.64	0.53	1.86	0.73	0.80	0
스트레스 조건											
25 th	0.56	0.04	-0.23	-0.10	-0.13	0.43	0.34	0.06	-0.25	-0.10	-0.14
Median	0.73	0.50	-0.04	0.02	-0.01	0.67	0.91	0.59	-0.07	0.02	-0.01
75 th	0.95	1.06	0.18	0.14	0.11	0.90	1	1.10	0.23	0.24	0.16

* 스트레스 조건과 통제 조건 간 차이가 유의미한 추정치 (two-tailed t -test, $p < .05$).

† $\Delta_+^{mb} - \Delta_-^{mb}$ 의 조건 간 차이 유의미 (two-tailed t -test, $p < .05$).

또한 옵션의 이전 기대치에 대한 가중치인 모수 γ 은 스트레스 조건의 피험자들에서 통제 조건보다 높았다($t_{50} = -3.31$, $p = 0.002$).

γ 는 옵션의 기대치가 업데이트에 이전의 경험의 영향이 반영되는 정도를 나타내며, 기본적인 TD 강화학습 모형의 학습률 learning rate α 를 1에서 뺀 값과 같다($\gamma = 1 - \alpha$). 따라서 높은 γ 추정치는 낮은 학습률 learning rate 을 의미하며, 상기 결과는 스트레스가 행동에 새로운 정보를 반영시키는 특성, 즉 학습률 learning rate 을 저하시켰음을 뜻한다.

RL model 3. RL model 2 를 적용한 추정 결과로 학습률 learning rate 에 대한 스트레스의 영향이 유의하게 나타난 것을 고려하며, 학습률 learning rate 에 있어서 모형 기반 model-based 의 요소와 모형 부재 model-free 요소를 구분했다. 즉, RL model 3에서는 γ 를 과제 구조(단계 전환 형태)에 따라 두 종류(70%의 확률로 단계가 전환된 상황에서의 γ_{common} 과 30%의 확률로 단계가 전환된 상황에서의 γ_{rare})로 구분되며, γ_{rare} 이 γ_{common} 에 비해 모형 기반 model-based 의 속성이 클 것을 가정한다. 이 모형의 모수 추정 parameter estimation 결과는 RL model 2 의 추정 결과와 유사하게 나타났다(표 1-3). 즉 RL model 3 의 모수 최대우도추정 결과, 스트레스 조건은 통제 조건에 비해 모형 부재 model-free 행동 경향이 크고 모형 기반 model-based 행동 경향이 적은 것으로 나타났다. 구체적으로 기술하면, 스트레스 조건의 피험자들은 통제 조건에 비해 Δ_+^{mb} 추정치가 낮게 나타난 동시에(양방향 t -검증 결과, $t_{50} = 2.43$, $p = 0.019$), Δ_-^{mb} 값은 덜 부적으로 나타났고($t_{50} = -3.13$, $p = 0.003$), 그 결과, 두 모형 기반 model-based 델타 값 간의 차이인 $\Delta_+^{\text{mb}} - \Delta_-^{\text{mb}}$

추정치가 통제 조건에 비해 낮았다($t_{50} = 2.89, p = 0.006$). 또한 스트레스 조건에서는 보상 받지 못했을 때 모형 부재 model-free 방식으로 행동을 전환시키는 경향에 대한 모수 Δ_{-}^{mf} 의 추정치가 더 부적인 값으로 추정되었다 ($t_{50} = 3.13, p = 0.003$). 한편 낮은 확률로 단계가 전환된 상황에서의 학습률 learning rate 이 반영되는 γ_{rare} 의 추정치는 스트레스 조건에서 통제 조건에 비해 높게 나타났으나($t_{50} = -3.02, p = 0.004$), 높은 확률로 단계가 전환된 상황에서의 γ_{common} 은 실험 조건 간 차이의 유의하지 않았다($t_{50} = -2.00, p = 0.052$). 이러한 결과는 스트레스가 모형 기반의 수행이 필요한 상황에서의 학습률 learning rate 을 저하시켰음을 시사한다.

다음으로는 모형 기반 model-based 의 델타 추정치와 모형 부재 model-free 델타 추정치에서 스트레스 효과가 유의하게 나타난 것이, 학습률 learning rate 에 대한 스트레스의 영향에 의한 것은 아닌지 확인하고자 했다². 각 델타 추정치에서 학습률 learning rate 의 영향을 통제하기 위해, (1) 각 델타 추정치($\Delta_{+}^{\text{mf}}, \Delta_{-}^{\text{mf}}, \Delta_{+}^{\text{mb}}, \& \Delta_{-}^{\text{mb}}$), (2) 강화 받은 시행에서의 델타 추정치, (3) 강화 받지 못한 시행에서 델타 추정치 간의 차이 값($\Delta_{+}^{\text{mf}} - \Delta_{-}^{\text{mf}}$ & $\Delta_{+}^{\text{mb}} - \Delta_{-}^{\text{mb}}$)들에 대해 두 종류의 학습률 learning rate ($1 - \gamma_{\text{common}}$ 와 $1 - \gamma_{\text{rare}}$)과 스트레스 처치 여부를 회귀변인으로 하는 회귀모형을 만들어서 회귀분석하였다.

² RL model 2 & 3에서의 델타 값들은 행동 결과에 따른 변화를 반영하는 모수와 학습률의 곱의 형태인 모수이기 때문이다.

표 1-4. RL model 3 의 델타 추정치들에 대한 회귀분석 결과

회귀 계수	추정치 평균 (표준 오차)	P 값
$\Delta_+^{mf} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	-0.35 (0.70)	0.613
b2	0.04 (0.56)	0.938
b3	-0.52 (0.38)	0.175
$\Delta_-^{mf} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	0.25 (0.26)	0.340
b2	-0.22 (0.21)	0.292
b3	-0.37 (0.141)	0.011
$\Delta_+^{mb} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	-0.51 (0.51)	0.318
b2	-0.15 (0.41)	0.720
b3	-0.50 (0.28)	0.084
$\Delta_-^{mb} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	-0.37 (0.25)	0.142
b2	0.35 (0.20)	0.086
b3	0.27 (0.14)	0.051
$\Delta_+^{mf} - \Delta_-^{mf} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	- 0.60 (0.72)	0.405
b2	0.27 (0.58)	0.648
b3	-0.15 (0.39)	0.704
$\Delta_+^{mb} - \Delta_-^{mb} = b_0 + b_1 \cdot (1 - \gamma_{common}) + b_2 \cdot (1 - \gamma_{rare}) + b_3 \cdot (stress)$		
b1	- 0.14 (0.65)	0.827
b2	- 0.50 (0.52)	0.346
b3	-0.76 (0.36)	0.037

분석 결과, 의사결정 시 모형 기반 model-based 행동 경향(Δ_+^{mb} - Δ_-^{mb})과 보상이 없을 때의 모형 부재 model-free 행동 경향(Δ_+^{mf} - Δ_-^{mf})에 대한 스트레스의 효과는 학습률 learning rate 의 영향을 통제한 후에도 유의한

것으로 나타났다(표 1-4). 이러한 결과들을 종합해볼 때, 스트레스는 부정적 결과 후의 모형 부재 model-free 행동 경향을 증진시키고, 모형 기반 model-based 행동 경향을 감소시키며, 모형 기반의 model-based 행동이 더욱 요구되는 상황에서 새로운 경험을 행동 선택에 반영하는 정도를 저하시키는(또는 이전 경험에 의존하는 정도를 증가시키는) 것으로 보인다.

3. 논의 1

연구 1에서는 스트레스가 습관적 행동을 촉진하는 영향에 기저하는 다면적인 요인을 탐색하고자 했다. 이를 위해 행동과 결과 간의 관계가 역동적으로 바뀌는 확률적 다단계 반전 학습 과제 probabilistic multi-stage reversal learning task 를 개발하여, 피험자들의 행동에 계산적 강화학습 모형 computational reinforcement learning model 을 적용했다. 실험 결과, 스트레스 처치 조건의 피험자들은 통제 조건에 비해 반전 학습 reversal learning 이 잘 이루어지지 않고 과제 수행 결과 총점이 낮은 등, 의사결정 과제의 전반적인 수행이 저조한 것으로 나타났다. 피험자들의 수행을 강화학습 모형 reinforcement learning model 에 적용하여 이론적 모수들의 추정치 parameter estimators 를 분석한 결과, 스트레스 조건 피험자들의 의사결정 행동은 통제 조건과 다음의 세 가지 차이가 있는 것으로 확인되었다. 첫째,

스트레스 조건은 모형 부재 model-free 행동, 특히 이전 행동의 결과가 강화 받지 못한 경우 다음 시행에서 모형 부재 model-free 방식으로 행동을 전환하는 양상이 통제 조건에 비해 높았다. 둘째, 스트레스 처치 받은 피험자들에게서는 모형 기반 model-based 행동 경향이 통제 조건에 비해 저조했다. 셋째, 스트레스 조건은 행동 선택 과정 중 새로운 정보를 받아들여서 다음 행동 선택에 반영하는 정도, 즉 학습률 learning rate 이 통제 조건에 비해 낮았다. 이러한 결과를 종합해보면, 스트레스는 행동의 습관 체계와 목표지향 체계의 영향에 모두 개입할 뿐 아니라 새로운 경험을 이용하는 정도를 저하시킴으로써 행동 선택에 부정적인 영향을 미칠 가능성을 시사한다.

최근 일부 실험 연구에서는 스트레스가 습관 행동을 촉진하는 현상이, 목표지향 행동의 희생으로 나타나는 것임이 주장된 바 있다. Schwabe 등(Schwabe et al., 2012)은 정상 성인들에게 체내 스트레스 호르몬 수준을 증가시키는 약물(hydrocortisone, yohimbine)을 복용시킨 후, 재평가 실험 패러다임을 사용한 이전 연구(Schwabe and Wolf, 2009)에서 스트레스가 습관행동을 촉진했던 것을 기능적 뇌영상 실험으로 재현했다. 연구 결과, 약물 복용 조건에서 결과 가치 감소에 대한 전전두엽 prefrontal cortex 일부 영역(orbitofrontal cortex, medial prefrontal cortex)의 신경활성화가 감소되었으며, 습관 행동의 주요 대뇌 영역으로 알려져 있는 복측 선조체 ventral striatum 영역에서는 스트레스 반응 관련 신경활성화 변화가 관찰되지 않았다. 연구자들은 이 결과를

바탕으로, 스트레스로 인한 습관 행동 선택은 목표지향적 과정에 개입하는 뇌 영역의 신경활성화 감소가 그 원인일 가능성을 제시했다. 그러나 재평가 과제 수행 중 스트레스로 인해 습관행동 영역으로 알려진 뇌 영역의 신경활성화 변화가 유의하지 않았다는 것이 Schwabe 등(2011)의 주장처럼, 스트레스가 습관 행동 처리과정을 변화시키지는 않는 것으로 간주하기는 어렵다. 최근 복측 선조체 *ventral striatum* 에서 습관 행동적 처리과정을 반영하는 신경신호 뿐 아니라, 목표지향적 행동 처리과정을 반영하는 신경신호가 함께 관찰된 바 있어, 두 행동 조절 체계의 신경활동 회로가 서로 배타적으로 구분되지 않을 가능성이 제기된 바 있다(Daw et al., 2011). 또한 다른 연구에서는 두 조절 체계의 활동을 반영하는 계산적 모형을 인간의 의사결정 행동에 적용한 결과, 전전두엽의 일부 영역(*inferior lateral prefrontal cortex*, *frontopolar cortex*)에서 두 체계의 활동을 비교하고 조절하는 양상이 관찰되었고, 이는 두 행동 체계가 전전두엽과 복측 선조체의 구별적인 신경활동 증가 혹은 감소로 인해 조절된다기 보다는 두 활동을 조절하는 역할을 담당하는 뇌 영역의 개입에 의해 조절될 가능성을 시사한다(Lee et al., 2014). 또한 Schwabe 등(2012)은 행동 선택에 대한 스트레스 효과를 탐색함에 있어서 행동의 강화물로 음식을 사용했으나 스트레스 호르몬 수준과 섭식 행동 간의 상호작용을 통제하지 못했으므로, 행동의 결과 가치를 조작한 시도가 스트레스 조건의 피험자들의 내적 상태를 변화시켰다는 것을 확신하기 어렵다.

한편 Otto 등은 정상인들에게 냉 자극 압박 실험 및 통제 실험 후 2 단계 Markov 의사결정 과제를 실시하게 했고, 그 결과 전체 피험자들의 실험(통제 실험과 스트레스 실험) 전후 코티솔 cortisol 의 변화 정도가 모형 기반 model-based 모수 추정치와 부적으로 상관관계가 유의했으며, 모형 부재 model-free 모수 추정치와는 상관관계가 유의하지 않았다(Otto et al., 2013b). Otto 등은 이러한 결과를 바탕으로, 스트레스로 인한 습관 행동 선택이 목표지향 체계의 조절 활동 장애로 일어나는 상보적인 현상임을 주장했다. 그러나 Otto 등의 연구에서는 피험자 전체에서 코티솔 변화량과 모형 기반 model-based 행동 모수 추정치 parameter estimator 간의 상관관계를 확인했을 뿐, 스트레스 처치 조건과 통제 조건 간 유의한 차이를 검증하지는 못했으므로, 두 행동 조절 체계에 미치는 스트레스의 영향을 밝혔다고 보기 어렵다. 따라서 스트레스로 인한 습관 행동 증가가 목표지향적 체계의 영향력 감소의 보상적인 현상이라는 주장은 아직 타당한 증거가 부족하며, 스트레스가 두 체계의 행동 조절 과정에 미치는 영향은 아직 명확하지 않다.

본 연구는 스트레스가 목표지향적 행동 조절 체계와 습관적 행동 조절 체계에 독립적인 영향을 미쳐, 두 체계의 행동 처리 과정을 모두 변화시킬 가능성에 대한 실험적 증거를 제시한다. 특히 모형 부재 model-free 행동의 정도를 두 종류, 즉 보상 받은 후 그 행동을 유지 win-stay 하는 경향성과 보상을 받지 못한 후 그 행동을 유지하지 않고 다른 행동으로 전환 lose-shift 하는 경향성으로 구분하여 추정하여, 스트레스가 전자를

변화시키지 않고 후자만을 선택적으로 변화시켰음을 확인했다. 스트레스 효과가 유의한 것으로 나타난 두 행동 모수추정치들은 서로 상관관계가 유의하지 않아, 스트레스 관련 변화가 서로 상보적이지 않음을 시사한다. 즉 본 연구는 스트레스가 습관 행동을 촉진하는 현상에는 (모형 기반 model-based 의사결정 과정의 손상과는 별도로) 모형 부재 model-free 행동 처리 과정 증진이 기여함을 제시하며, 모형 부재 model-free 행동에 대한 스트레스의 영향은 강화학습 및 의사결정 과정 중 결과 종류(보상 혹은 무보상)에 따라 달라질 수 있음을 제시한다. 뇌의 도파민성 신경세포들은 행동의 부정적인 결과를 처리할 때, 긍정적인 결과 처리와 구분되는 신경활성화를 보인다(Jensen et al., 2007; Bromberg-Martin et al., 2010)는 것을 고려하면, 스트레스가 의사결정 과정에서 부적 결과에 대한 기저 뇌신경활동과 정적 결과에 대한 신경활동 양상에 서로 다른 조절 작용을 할 가능성이 있으며, 이에 대해서는 추후 기능적 뇌영상 연구에서 확인할 필요가 있다.

한편 본 연구 결과는 스트레스로 인한 습관 행동 선택 증가가 모형 기반 행동 감소 때문만이 아니라 변화하는 상황에서 매번 새로운 정보를 습득하여 행동에 반영하는 학습률 learning rate 이 저하된 탓일 가능성을 제시한다. 본 연구에서는 의사결정자가 과제 구조에 대한 지식, 즉 단계 전환 확률을 고려하는 모형 기반 model-based 행동과, 과제의 매 시행마다 경험하는 행동-결과 수반성 contingency 에 대한 새로운 정보를 다음 시행의 행동에 이용하는 학습률 learning rate 을 구분했다. 모수 추정치 parameter

estimator 에 대한 분석 결과, 스트레스는 모형 기반 model-based 행동 경향뿐 아니라 학습률 learning rate 도 감소시킨 것으로 나타났다. 특히 학습률 learning rate 을 의사결정 상황에 따라 두 종류로 구분하여 추정한 결과, 모형 기반 model-based 행동이 더욱 요구되는 상황(30%의 단계 전환 후 의사결정 상황)에서의 학습률 learning rate 에서는 스트레스 처치 효과가 유의했던 것과 달리, 그렇지 않은 상황(70%의 단계 전환 후 의사결정 상황)에서의 학습률 learning rate 에서는 실험 조건 간 차이가 유의하지 않았다. 강화학습 모형에서 학습률 learning rate 은 선택 행동에 대한 기대가치와 실제 가치 간의 차이, 즉 보상에측오류에 의해 행동의 기대가치 expected value 가 얼마나 잘 업데이트되는지를 반영하는 모수 parameter 이다(Sutton and Barto, 1998). 매 시행마다의 예측오류 prediction error 는 매번 상황에 따라 얻게 되는 새로운 정보의 역할을 하여, 학습률 learning rate 에 의해 그 영향력의 크기가 조정되어 이후의 행동 선택에 반영된다(Behrens et al., 2007; Bernacchia et al., 2011; Simon and Daw, 2011). 적응적인 의사결정을 위해서는 새로운 정보를 얻었을 때 그것이 효과적으로 행동에 반영되어야 하며, 환경 변화에 직면했을 때 새로운 정보가 활용되지 않은 채 이전 행동이 지속되면 역동적인 환경에서 의사결정을 하기 어렵다. 따라서 본 연구 결과는 행동-결과 관계가 역동적으로 변화하는 반전 학습 reversal learning 상황 중에서도 모형 기반의 행동이 요구되는 상태에 처했을 때, 스트레스가 유연하고 적응적인 행동 선택을 방해할 가능성을 시사한다.

요약하면, 연구 1에서는 스트레스가 강화학습의 다양한 요소에 부정적인 영향을 미침으로써 의사결정 행동에 관여함을 밝혔다. 즉, 스트레스는 모형 기반 강화학습 (model-based reinforcement learning) 을 저하시키고 모형 부재 강화학습 (model-free reinforcement learning) 을 증진시키며, 새로운 정보를 다음 행동 선택에 반영하는 비율인 학습률 (learning rate) 을 저하시키는 것으로 나타났으며, 본 연구 결과는 스트레스가 부적응적인 의사결정을 초래하는 인지행동적 기전을 제시한다.

Ⅲ. 연구 2

연구 2에서는 스트레스가 행동 선택에 미치는 영향이 신경활동 수준에서는 어떻게 나타나는지에 대해 기능적 뇌 영상 실험 기법을 이용하여 탐색했다. 또한 스트레스가 습관적 행동을 촉진하고 목표지향적 행동을 방해하는 현상이 스트레스의 정도와 무관하게 일관적으로 나타나는 것인지, 아니면 스트레스 경중에 따라 그 효과가 다르게 나타나는지 알아보고자 했다. 스트레스는 그 정도가 심각하거나 지속적으로 반복될 경우 인지 기능 과제 수행을 저하시키지만, 일시적이거나 경미한 스트레스는 인지 기능 발휘에 긍정적인 효과를 초래한다는 증거들이 있다(Gold et al., 1975; Sauro et al., 2003; Domes et al., 2005; Roozendaal et al., 2009; Salehi et al., 2010). 또한 각성 수준과 인지 기능 수행 간의 관계를 역 U 자 형태로 정의하는 여키스-도슨 법칙 Yerkes-Dodson law (Yerkes and Dodson, 1908)을 고려하면, 연구 1에서 확인된 스트레스가 의사결정에 미치는 부정적인 영향이 처치 수준에 따라 늘 같은 양상으로 나타날 지가 의문이다. 현재로서는 스트레스와 행동 선택의 관계를 탐색한 선행 연구들 결과, 스트레스가 습관적인 행동을 촉진한다는 것이 제시되고 있을 뿐, 스트레스의 영향이 처치 수준에 따라 달라지는지 여부에 대해서는 아직 확인된 바가 없다. 이에 연구 2에서는 정상 성인들을 스트레스 무처치 조건 no-stress-treatment, 단일처치 single-stress-treatment 조건, 이중처치 double-stress-treatment 조건에

할당하여 각기 다른 수준의 스트레스를 처치함으로써, 의사결정 수행에 대한 스트레스의 정도에 따른 영향을 인지행동적으로 탐색했을 뿐 아니라 기능적 뇌 영상 실험을 통해 뇌신경활동 수준에서도 탐색했다.

1. 연구 방법

1) 연구대상

서울대학교에서 심리학 관련 수업을 수강하는 학부생들 중 연구 참여를 희망했던 자원자들과 학교 인터넷 게시판에 게시한 연구참여자 모집 광고를 통해 신청한 학생들을 대상으로 참여 조건에 맞는 정상 성인을 선별했다. 연구 참여자 선별 조건은 연구 1과 동일하며, 기능적 자기공명영상 촬영 시 자기장 내에서 안전 상의 문제가 되거나 BOLD 신호에 문제가 발생할 만한 사항들(체내에 금속 물질이나 전자 장치가 삽입된 경우, 폐소 공포증이 있는 경우, 의치나 치과교정기가 삽입되어 있는 경우 등)에 대해 사전 확인을 거쳐 참가를 제한하였다. 총 64명의 연구대상자들을 스트레스 무처치_{no-stress-treatment}, 스트레스 단일처치_{single-stress-treatment}, 스트레스 이중처치_{double-stress-treatment} 조건에 각각 무선 할당하여, 실험 전 6시간 동안은 자율신경계 각성이 유발되는 카페인 음료 마시기, 심한 신체 운동을 삼가 하도록 한 후, 스트레스 처치 실험

및 기능적 뇌 영상 실험을 진행했다.

64명의 기능적 자기공명영상 실험 결과, 영상 촬영 중 4mm이상 머리를 움직인 것으로 나타난 피험자, 의사결정 과제 수행 시 한 옵션을 95% 이상 지속적으로 선택하여 학습이 잘 이루어지지 않았다고 판단된 피험자, 자기공명영상 실험 중 신체적 불편함을 호소하여 실험을 중단한 피험자, 졸음 등으로 인해 과제 중 10% 이상의 시행에서 반응이 누락된 피험자들을 제외하고, 총 38명(스트레스 무처치 13명, 스트레스 단일처치 13명, 스트레스 이중처치 12명)의 실험 결과를 분석했다.

모든 연구대상자들은 연구참여에 앞서 연구의 전반적인 절차에 대해 안내를 받았으며, 서면 동의서에 서명함으로써 연구참여에 대한 자발적인 의사를 밝혔다.

2) 스트레스 유발 실험

연구 2는 기능적 뇌 영상 실험으로 진행해야 하므로, 실험 환경이 연구 1의 행동 실험 환경과 동일하지 않다. 즉 피험자들은 스트레스 처치를 받은 후 뇌 영상 촬영 장치 안에 누운 상태에서 약 5~6분 간의 해부학적 영상 촬영 후 30여분 간의 본 실험 과제를 수행하게 된다. 따라서 연구 2에서의 과제 수행 상황은 스트레스 반응을 보다 빠르게 안정화시키거나 각성 상태를 저하시킬 가능성이 있으며, 이와 관련하여

사회적 평가 상황에서의 냉 자극 압박 실험^{socially evaluated cold pressor test, SECPT} 처치는 그 효과가 연구 1에서와 달리 경미할 가능성이 예상된다. 이에 연구 2에서는 단일 스트레스 처치 조건에 해당되는 연구참여자들에게서 경미한 수준의 스트레스 효과를 얻기 위해 연구 1에서 사용한 SECPT를 실시했고, 이중 스트레스 처치 조건의 피험자들에게는 SECPT 실시 후 추가적으로 두 번째 스트레스 처치를 했다. 즉 이중 스트레스 처치 조건의 피험자들은 SECPT 처치 후 자기공명 뇌 영상 장치에 들어가서 누운 채로 5분 50초간 동영상 시청하게 했는데, 처치에 사용된 동영상은 ‘김복남 살인사건의 전말(Bedeviled, 2010, Cheol-Soo Jang)’이라는 영화 중 일부분을 편집한 것으로, 남녀간의 잔인한 폭력 장면을 담고 있다. 동영상 상영 중에는 영화의 내용에 따라 고통으로 울부짖거나 분노로 고함을 지르는 등의 음향을 헤드폰을 통해 듣게 했다. 실험 전 피험자들에게 냉 자극 압박 처치나 영화 시청으로 인해 심리적인 불쾌감을 경험할 수 있음을 사전에 설명했으며, 실험 도중 원한다면 언제든지 처치를 중단할 수 있음을 안내했다. 스트레스 무처치 조건의 피험자들에게는 연구 1의 비교 통제 조건과 같은 과정을 거치게 했다(36 ~ 38°C의 미지근한 물에 한 손을 손목까지 1분 간 담그게 하고, 그 과정은 녹화하지 않음). 세 실험 조건의 처치 효과를 확인하기 위해 실험 중 ‘전혀 스트레스를 느끼지 않았음’을 ‘0’, ‘매우 큰 스트레스를 느꼈음’을 ‘100’으로 시각화한 시각적 아날로그 척도^{Visual Analogue Scale, VAS}를 사용했다. 모든 실험은 스트레스 호르몬인 코티솔의

일주기 리듬이 안정화되는 오후 시간인 1시부터 5시 사이에 실시했다. 스트레스 무처치 조건과 단일 스트레스 처치 조건의 피험자들은 SECPT 및 미지근한 물에 손 담그기 실시 10분 후 기능적 뇌 영상 촬영 중 의사결정 과제를 수행하도록 했고, 이중 스트레스 처치 조건의 피험자들은 두 번째 스트레스 유발 실험인 스트레스 영상 기법 실시 10분 후에 기능적 뇌 영상 촬영 중 의사결정 과제를 수행하도록 했다.

3) 의사결정 과제

연구 1에서 사용한 두 단계의 반전 학습 의사결정 과제_{2-stage reversal learning decision task}를, 시행_{trial} 수가 총 192개인 과제로 수정하여 사용했다(그림 2-1). 즉 구간_{block} 1 ~ 4는 연구 1에서와 같이 40 시행이 모두 포함되며, 구간_{block} 5는 32 시행만 포함되고, 구간_{block} 6은 생략되었다. 과제의 총 192개의 시행은 4 개의 런_{run}으로 나뉘었고 각 런_{run} 마다 48 시행이 진행되었다. 과제는 사건 관련 설계_{event-related design}로 계획되어, 한 시행 당 첫 번째 의사결정, 두 번째 의사결정, 결과 확인의 총 세 종류의 사건이 발생한다(그림 2-1B). 두 의사결정 단계는 지속 시간을 각각 2초로 고정시켜서_{time-locking} 고정된 시간 내에서 피험자가 자기 속도로 옵션을 선택하도록_{self-paced} 했고, 결과 확인 단계는 1초로 고정시켰다. 한 시행에서 다음 시행으로 넘어갈 때의

시행 간격은 무선적으로 1초 또는 3초 또는 7초로 지정하여 실험 결과로 얻은 BOLD 신호에 대한 혈액학 반응 함수(hemodynamic response function)에서의 모수 추정치 계산의 효율성을 높였다.

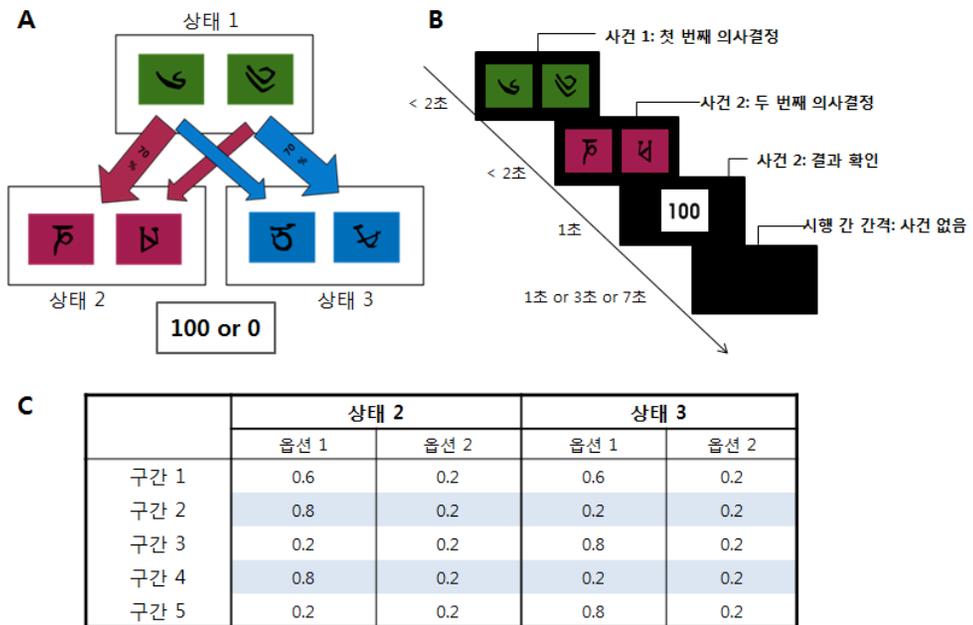


그림 2-1. fMRI 실험 과제 및 과제 중 한 시행에 대한 타임라인

4) 기능적 자기공명영상 획득

기능적 뇌 영상 실험은 서울대학교 뇌 영상 센터에서 진행했고, 3 Tesla 자기공명 영상 장비(MAGNETOM Trio; Siemens, Germany)를 이용하여 뇌 영상을 획득했다. 기계에 대한 소음을 줄이고자 자기공명영상 장치에 들어가기에 앞서 연구참여자들에게 일회용 귀마개를 착용시켰고, 바로 누운 상태에서 헤드 코일에 머리를 고정한 후 움직임을 최소화하기 위한 고정 스폰지를 끼워 넣었다.

기능적 뇌 영상 실험 전에 전체 뇌에 대한 T1 강조 경사사장을 이용하여 해부학적 구조 영상을 다음과 같은 방식으로 촬영했다: 반복 시간(repetition time, TR) = 1900ms, 에코 시간_{echo-time}, TE = 2.36, 영상 범위_{field of view}, FOV = 256 x 256mm, 숙임각_{flip angle}, FA = 9°, 부피소 크기_{voxel size} = 1 x 1 x 1mm. T1 강조 영상 상에서 뇌의 구조적 이상 유무를 확인하고 기능적 뇌 영상 촬영 시 대뇌 전체가 포함되도록 x, y, z축 상의 범위를 조정했다. 기능적 자기공명영상 자료는 T2*-강조 경사사장 echo-planer imaging (EPI) 펄스 열을 사용했으며, 촬영 방식은 다음과 같다: TR = 2000ms, TE = 30ms, FOV = 240 x 240, FA = 79°, 부피소 크기_{voxel size} = 3 x 3 x 3mm, 횡단면 간격_{slice gap} = 1mm, 횡단면 수_{number of slices} = 30.

실험 과제 수행 시 EPI 영상은 각 런_{run} 별로 210개의 영상을 획득했고, 4개의 런_{run}을 통해 피험자 별로 820개의 영상을 획득했다. 기능적 뇌 영상 획득 시, 각 회기마다 자기적 평형 상태에 도달하기 전의 처음 6초 간은 신호만 들어가고 영상은 저장하지 않는 더미

스캔dummy scan이 포함되었다.

5) 실험 절차

연구참여자로 선정된 피험자들은 선별 검사 후 한 달 이내에 실험실을 방문하여 스트레스 유발 실험과 의사결정 과제 수행을 하도록 했다. 모든 실험은 스트레스 호르몬인 코티솔cortisol의 일주기 리듬diurnal rhythm이 안정화되는 오후 시간인 1시에서 5시 사이에 진행했다. 피험자들은 뇌 영상 센터 대기실에 입실하여 두 단계 반전 학습 의사결정 과제에 대한 설명을 듣고 나서, 냉 자극 압박 실험SECPT 혹은 통제 실험 절차에 따라 처치를 받았다. 처치가 끝난 피험자는 곧바로 자기공명 뇌 영상 장치가 있는 방으로 이동하여 뇌 영상 실험에 대해 안내 받았다. 뇌 영상 촬영에 대한 안내가 끝나면, 스트레스 무처치 조건과 단일 스트레스 처치 조건의 피험자들은 해부학적 구조 영상을 촬영을 시작하여, 촬영 중 의사결정 과제에 대한 연습 세션을 수행한 후, 본 실험 과제를 실시했다. 스트레스 무처치 조건과 스트레스 단일처치 조건의 피험자들이 스트레스 처치 (혹은 통제 처치) 후 본 실험 과제를 실시하는 시점까지는 10 ~ 15분의 시간적 간격을 두었다. 반면 스트레스 이중처치 조건의 피험자들의 경우(그림 2-2), 대기실에서 일차적인 스트레스 처치를 받은 후 뇌 영상 촬영실로 이동하여 뇌 영상

촬영에 대한 안내를 받고, 곧바로 뇌 영상 장치에 누워 그 안에서 이차적인 스트레스를 처치를 받았다. 이차 스트레스 처치가 끝나면 해부학적 구조 영상 촬영을 시작하여 촬영 중 의사결정 과제에 대한 연습 세션을 진행했고, 연습 세션이 끝나면 본 실험 과제를 실시했다. 이중 스트레스 처치 조건의 경우 본 실험 과제 시작 시점과 이차 스트레스 처치 간 시간 간격은 10분을 유지했다(본 실험 과제 시작과 일차 스트레스 처치 간 시간 간격은 20분에 해당됨).



20

그림 2-2. 스트레스 이중처치 조건에 대한 실험 절차

6) 뇌 영상 정보의 전처리

기능적 자기공명 뇌 영상 자료의 전처리_{preprocessing}와 분석은 SPM 8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK)을 사용했다. 다음과 같은 절차로 전처리했다: 1) fMRI 촬영 시 발생한 연구참여자들의 머리 움직임을 6개의 모수를 통한 정합_{3-D rigid body registration} 과정으로 보정; 2) 자기공명영상 획득 시 발생하는 횡단면 영상 간의 시간 차를 보정; 3) 각 참여자의 기능적 영상은 표준화된 Montreal Neurological Institute(MNI) 공간으로 정상화하는 방법으로 공간적 표준화 작업_{spatial normalization}을 시행. 통계적 모수 지도화 분석을 위한 자료의 전처리는 3 x 3 x 3mm 크기의 부피소로 재추출했고, 표준 공간으로 정상화한 기능적 영상은 모두 6mm FWHM Gaussian kernel로 편평화_{smoothing} 했다.

7) 행동 결과 분석

세 실험 조건 간 스트레스 처치로 인한 주관적 스트레스 정도 및 과제 수행 차이를 검증하기 위해 일원분산분석_{ANOVA}을 했다. 분석에 사용한 각 피험자의 행동 측정치들은 다음과 같다. 1) 첫 번째 단계에서 두 옵션_{option} 중 하나를 고를 때까지 걸린 반응 시간의 평균, 2) 전체 수행 결과로 누적된 점수, 3) 첫 번째 단계_{stage}에서 유리한 옵션(구간_{block} 2, 4에서는 ‘상태_{state} 2’와 70%로 연결된 옵션; 구간_{block} 3, 5에서는

‘상태_{state} 3’과 70%로 연결된 옵션)을 선택한 확률.

피험자들의 반전 학습_{reversal learning}의 수행에 대한 스트레스 효과를 검증하기 위해, 구간 3 ~ 6의 유리한 옵션 선택 확률에 대한 혼합 설계 분산분석_{mixed-design ANOVA}을 실시했다. 피험자 내 요인으로 각 구간 내의 하위 구간_{sub-block} 종류(전반부_{initial half}: 시행 1 ~ 20; 후반부_{final half}: 시행 21 ~ 40)를 설정했고, 피험자 간 요인으로는 스트레스 처치 종류(무처치, 단일처치, 이중처치)를 설정했다.

다음으로는 피험자들의 첫 번째 단계에서의 의사결정 행동 유지 확률에 대한 요인을 세 가지로 설정하여, 3요인 혼합 설계 분산 분석_{mixed design ANOVA}을 실시했다. 피험자 내 요인 두 가지로 직전 수행의 보상 여부와 두 번째 단계가 첫 번째 선택과 강한 확률 관계인지 약한 확률 관계인지 여부를 설정했고, 피험자 간 요인으로는 스트레스 처치 종류(무처치, 단일처치, 이중처치)를 설정했다.

8) 강화학습의 계산 모형 적용

연구 1에서 적용된 세 종류의 강화학습 계산 모형 중 2 단계 반전 학습 의사결정 과제 수행 결과에 가장 잘 부합된다고 판단되는 RL model 2를 적용시켜서, 피험자 별로 다섯 종류의 모수(γ , Δ_+^{mf} , Δ_-^{mf} , Δ_+^{mb} , Δ_-^{mb})를 최대우도 추정 _{maximum likelihood estimation} 방법을 통해

추정했다. 모수의 최대우도 추정 maximum likelihood estimation 에는 Matlab (Mathworks, Inc) 기반의 알고리즘 fminsearchbnd 을 이용했고, γ 만 0 에서 1 사이의 범위 내에서 추정하고 다른 모수들은 추정 범위를 제한하지 않았다. 모수 추정 시에는 모수마다 무선적으로 선택된 초기값으로 사용하여 우도 함수 likelihood function 내에서 지역적인 최대값 추정 가능성 local maxima problem 을 피하고 함수 전체의 최대값을 구할 때까지 추정 과정을 반복했다.

9) 뇌 영상 자료 분석

의사결정 관련 뇌 활성화 양상이 스트레스 처치로 인해 어떻게 달라지는지 탐색하고자 했다. 우선 첫 번째로, 과제 수행 중, 목표지향적 행동 요인과 습관적 행동 요인이 모두 요구되는 첫 번째 단계의 행동 선택 시, 스트레스가 기저 신경활성화 양상의 실험 조건 간 차이를 분석했다. 이를 위해, 각 피험자마다 획득한 420개의 fMRI 영상으로부터 혈역학 반응 모형 hemodynamic response model으로 적합성의 유의도를 부피소 voxel 별로 산출했다. 즉 각 시행의 첫 번째 단계 stage 의사결정 사건 event, 두 번째 단계 의사결정 사건, 결과 제시 사건을 회귀변수 regressor로 설정하여, 다음과 같은 일반선형모형 general linear model. GLM을 통해 피험자 별로 각 사건 event의 회귀계수들을 추정하였다.

GLM 1: $Y = b_0 + b_1 * (\text{첫 번째 의사결정}) + b_2 * (\text{두 번째 의사결정}) + b_3 * (\text{결과 제시})$

GLM 1을 통해 피험자 개인 수준에서 첫 번째 의사결정 사건의 회귀계수를 추정하여 다음과 같이 두 가지 분석을 실시했다.

- (1) 스트레스 무처치 조건과 단일처치 조건 간 차이를 분석하여, 스트레스 단일처치로 인한 첫 번째 의사결정 사건 관련 뇌 활성화 양상의 변화를 탐색했다. 분석 결과의 역치는 $p_{\text{uncorrected}} < 0.001$, 군집 크기 $k > 50$ 로 삼았다.
- (2) 위 분석 결과 중 스트레스 단일처치 조건에서 무처치 조건보다 첫 번째 의사결정 사건 시 신경활성화가 크게 나타난 영역을 관심 영역(region of interest, ROI)으로 지정하여, 해당 영역들의 뇌신경활성화가 스트레스 이중처치 조건에서 저하되는지 여부를 ROI 분석으로 확인했다. 분석 결과는 지정한 ROI 내에서 Family-wise 방식으로 교정하여 $p_{\text{corrected}} < 0.05$ 수준을 역치로 삼았으며, 군집 크기의 역치는 $k > 50$ 으로 하였다.

다음은 GLM 1에서 첫 번째 의사결정 사건과 결과 제시 사건에 각각 매 시행의 선택 행동의 기대치(expected value of a chosen action, chosen value)와 예측 오류(prediction error)를 모수 조절 변수(parametric modulator)로 지정함으로써 두

가지 회귀변수_{regressor}를 추가한 모형 (GLM 2)을 설정했다.

$$\text{GLM 2: } Y = b_0 + b_1 * (\text{첫 번째 의사결정}) + b_2 * (\text{첫 번째 의사결정}) * (\text{선택 행동의 기대치}) + b_3 * (\text{두 번째 의사결정}) + b_4 * (\text{결과 제시}) + b_5 * (\text{결과 제시}) * (\text{예측 오류})$$

이 모형에서는 강화학습 모형인 RL model 2를 통해 매 시행마다 계산되는 선택 행동의 기대치_{chosen value}와 예측 오류_{prediction error}가 시간 경과_{time course}에 따른 해당 사건과 일치하는 정도가 회귀계수로 추정된다. 따라서 이러한 모형을 통해 선택 행동의 기대치_{chosen value} 및 예측 오류_{prediction error}를 반영하는 BOLD 신호가 유의한 영역을 탐색할 수 있다. 선택 행동의 기대치_{chosen value}에 대한 회귀변수와 예측 오류_{prediction error} 회귀변수는 모두 혈역학 반응 함수_{hemodynamic response function}로 조정_{convolve}하였다.

피험자 별 매 시행의 선택 행동의 기대치_{chosen value}와 예측 오류_{prediction error} 계산 시에는 각 조건에서 모수 추정치를 이용했다. 즉 세 실험 조건에서 추정된 세 세트_{set}의 모수 추정치들을 사용하여, 각 조건에서의 피험자들의 선택 행동의 기대치_{chosen value}와 예측 오류_{prediction error}를 계산했다. 최대우도 추정치들은 정규화된 값들이 아니므로, 개인별 모수 추정치 뇌 영상 자료 분석에 있어서 잡음_{noise}이 크게 나타날 수 있다. 따라서 fMRI 자료에 사용되는 강화모형 모수치는 조건 내의 고정효과_{fixed effect}를 가정함으로써 정규화_{regularization}의 형태

모수추정치를 구할 수 있어, 선행 연구들에서 이 방법을 사용한 바 있다. (O'Doherty et al., 2004; Daw et al., 2006; Schönberg et al., 2007). 따라서 선택 행동의 기대치_{chosen value}와 예측 오류_{prediction error}를 각 실험 조건 내에서 고정효과를 갖는 RL model 2 모수 추정치들로 계산한 후, GLM 2의 회귀변수로 사용했으며, 해당 회귀계수를 피험자 개인 수준에서 추정한 후 다음과 같은 분석을 실시했다.

- (1) 선택 행동의 기대치 관련 뇌 활성화 영역과 예측 오류 관련 뇌신경 활성화 영역을 각 실험 조건 별로 탐색했다. 즉 선택 행동의 기대치와 예측 오류에 대해 세 실험 조건 각각에서 한 집단 t -test를 실시했고, 분석의 역치는 $p_{\text{uncorrected}} < 0.001$, 군집 크기 $k > 100$ 으로 지정했다.
- (2) 선택 행동의 기대치, 예측 오류 관련 신경활성화의 조건 간 차이가 유의한 뇌 영역을 탐색했다. F -test를 실시했고, 분석의 역치는 $p_{\text{uncorrected}} < 0.001$, 군집 크기 $k > 20$ 으로 지정했다.

2. 연구 결과

1) 행동 분석 결과

의사결정 과제 수행 결과(표 2-1)에 대한 스트레스 처치 효과 및 스트레스 처치 종류(무처치, 단일처치, 이중처치)의 효과가 유의한지 알아보기 위해, 주관적 스트레스 자기보고 측정치(VAS 점수), 과제 수행 중 첫 번째 선택 시 평균 반응 시간, 수행 결과 총점, 구간 2 ~ 5 에서의 유리한 옵션 선택 비율에 대해 스트레스 처치 종류를 요인으로 하여 일원분산분석을 실시했다. 분석 결과, 세 조건의 주관적 스트레스 자기보고 측정치($F_{(2,35)} = 172.51, p < .001$), 반응시간($F_{(2,35)} = 4.21, p < .02$), 유리한 옵션 선택 확률($F_{(2,35)} = 3.31, p < .048$)에 대한 스트레스 처치 효과가 유의했고, 과제 수행 결과 총점에 대한 스트레스 처치 효과는 유의하지 않았다($F_{(2,35)} = 0.03, p < .973$).

표 2-1. 피험자들의 조건 별 과제 수행 결과 평균 (표준오차)

	스트레스 무처치	단일 스트레스처치	이중 스트레스처치
주관적 스트레스 (VAS)	0.69 (0.14)	5.93 (0.42)	12.37 (0.65)
반응시간	0.64 (0.027)	0.65 (0.030)	0.74(0.026)
과제 수행 결과 총점	9130.77 (338.05)	9238.46(468.71)	9125.00(321.965)
유리한 옵션 선택 확률	0.63 (0.02)	0.57 (0.03)	0.52 (0.03)

일원분산분석 결과에 대해 사후분석으로 Turkey's HSD 검증을 실시한 결과, 주관적 스트레스 자기보고 측정치는 유의확률 $p < .001$ 수준에서 세 조건 간 차이가 유의했다. 즉, 스트레스 이중처치

조건은 단일처치 조건에 비해 VAS 점수가 높았고, 단일처치 조건은 무처치 조건에 비해 VAS 점수가 유의하게 높았다. 이러한 결과는 스트레스 처치로 인해 피험자들이 주관적으로 스트레스를 느꼈음은 물론, 단일처치와 이중처치 간 차이가 주관적 스트레스의 정도에 반영되었음을 뜻한다.

첫 번째 단계 stage 에서의 의사결정의 평균 반응 시간에 있어서는 스트레스 이중처치 조건이 무처치 조건에 비해 유의하게 길었고($p = .027$), 유리한 옵션 선택 확률은 스트레스 이중처치 조건이 무처치 조건에 비해 유의하게 낮았다($p = .038$). 즉, 스트레스 이중처치는 무처치에 비해 의사결정에 걸리는 시간을 지연시켰고, 유리한 옵션을 선택할 가능성을 저하시킨 것으로 보인다.

반전 학습 양상 분석. 행동(특정 옵션을 선택하는 행동)과 결과 간의 수반성 contingency 이 역동적으로 변화하는 상황에서 보상을 받기 위해서는 변화에 따라 행동을 유연하게 바꾸어야 한다. 따라서 다음으로는, 스트레스가 변화에 대한 대처행동에 처치 수준에 따라 영향을 미치는지 확인하고자 했다. 본 과제는 구간 block 2 ~ 5 에서 첫 번째 단계의 두 옵션의 기대치가 구간마다 반전이 된다. 이에 구간 block 3 ~ 5 에서는 각 구간 내 전반부 initial half 와 후반부 final half 에서 유리한 옵션을 선택하는 확률의 역동적인 변화가 일어날 것이 예상된다. 따라서 구간 block 3 ~ 5 의 각 구간 block 내 하위 구간 sub-block 을 전반부와 후반부로 나누어 스트레스

처치에 따른 행동 선택 양상을 분석했다. 즉, 구간_{block} 3 ~ 5의 각 구간 내 전반부(20 개의 시행)와 후반부(나머지 시행)에서, 첫 번째 단계의 유리한 옵션(이를 테면 구간 3, 5에서는 옵션 2, 구간 4에서는 옵션 1)을 선택한 확률을 구분했다. 첫 번째 단계의 두 옵션 중 가치가 우월한 것을 선택한 확률에 대해, 피험자 내 요인으로 구간 내의 하위 구간 종류(전반부, 후반부)를 지정하고, 피험자 간 요인으로는 스트레스 처치 종류(무처치, 단일처치, 이중처치)를 지정하여 혼합 설계 분산분석 mixed design ANOVA 을 실시했다.

분석 결과(그림 2-3), 구간 내 하위 구간_{sub-block} 종류의 주효과가 유의했으며($F_{(1,35)} = 46.04, p < .001$) 이는 피험자들의 과제 수행 결과에 반전 학습 reversal learning 의 특성이 잘 나타났음을 뜻한다. 구간 내 하위 구간_{sub-block} 종류와 스트레스 처치 종류 간 상호작용 효과도 유의했으며($F_{(2,35)} = 5.61, p = .008$), 이는 스트레스가 피험자들의 행동과 결과 간 수반성 contingency 변화에 대한 적응적 행동 선택에 유의한 영향을 미쳤음을 시사한다.

추가적인 분석 결과, 구간 내 전반부에서의 유리한 옵션 선택 비율은 세 실험 조건 간 차이가 유의하지 않았다($F_{(2,35)} = 1.02, p > .371$). 그러나 후반부의 유리한 옵션 선택 비율은 세 조건 간 차이가

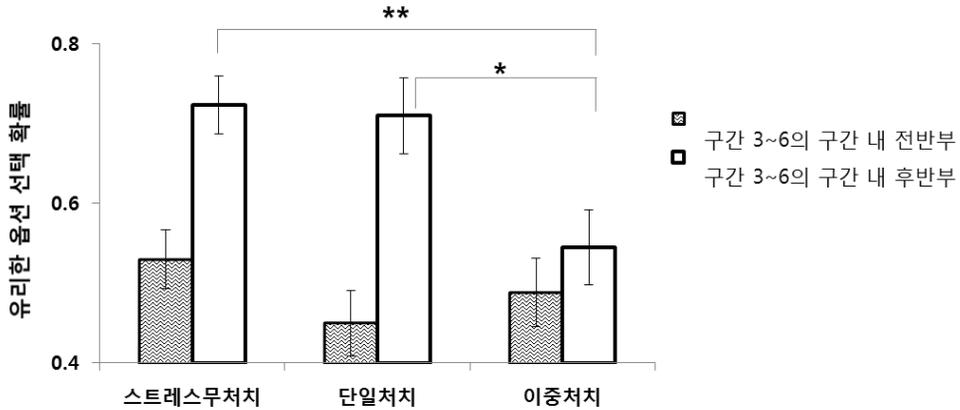


그림 2-3. 반전학습(reversal learning)에 대한 스트레스 처치 수준 별 효과. 구간 3 ~ 5에서 유리한 옵션을 선택하는 행동 비율의 구간 내 전반부와 후반부 간 차이와 스트레스 처치 간의 상호 작용 효과가 유의하게 나타남. 오류 막대는 평균의 표준 오차. * $p < .05$, ** $p < .01$

유의했다($F_{(2,35)} = 5.02, p = .012$). 이는 스트레스 이중처치 조건의 피험자들이 다른 두 조건에 비해 과제 수행 중 상황적인 변화, 즉 행동과 결과 간 수반성(contingency)의 변화에 유연하게 대처하지 못한 것에 기인하는 것으로 보인다. 즉, 스트레스 무처치 조건과 단일처치 조건 간 차이는 유의하지 않았으나($p = .904$), 스트레스 이중처치 조건은 다른 두 조건에 비해 반전 학습 구간 내 후반부에서 유리한 옵션을 선택한 확률이 유의하게 낮았다(무처치 조건과의 차이: $t_{23} = 2.93, p = .008$; 단일처치 조건과의 차이: $t_{23} = 2.52, p = .019$). 이는 스트레스 이중처치 조건에서는 수반성(contingency)의 반전(reversal)이 일어난 후 일정 양의 시행 동안 새로운 수반성(contingency)을 경험한 후에도 수행이 개선되지 않았음을

뜻하며, 이러한 양상은 스트레스 단일처치 조건에 비해서도 큰 것으로 보인다.

이전 행동 유지-전환 양상 분석. 스트레스가 의사결정 행동 과정 중 습관 행동 체계와 목표지향적 행동 체계의 영향력을 각각 어떻게 변화시켰는지 탐색하기 위해, 첫 번째 단계의 의사결정 시 매 시행마다 행동 유지하거나 전환하는 양상에 스트레스 처치 효과가 유의한지 분석했다. 모형 부재 model-free 행동 선택 체계, 즉 습관 체계만 사용하여 행동을 선택하는 의사결정자라면, 매 시행에서 이전 행동을 유지할지 전환할지 여부가 이전 시행의 선택이 보상을 받았는지 못 받았는지에 따라 정해질 것이다. 반면 모형 기반 model-based 행동 체계, 즉 목표지향적 체계만을 기반으로 행동을 선택하는 의사결정자라면 과제의 전체적인 구조를 파악하여 그 지식을 바탕으로 최적의 행동을 선택할 것이므로, 매 시행에서 의사결정을 할 때 이전 시행에서 자신이 선택한 옵션이 70%의 확률 관계인 화면으로 전환되었는지, 30%의 관계인 화면으로 전환되었는지를 고려할 것이다. 이에 피험자가 과제 첫 번째 단계에서 이전 행동을 다음에도 유지할 확률에 대해 두 가지 피험자 내 요인, 즉 이전 행동의 결과인 보상 여부와 과제 구조, 즉 이전 행동으로 인한 단계 전환 형태(70%의 전환 또는 30%의 전환)를 가정했다. 또한 피험자 간 요인으로서 스트레스 처치 조건의 종류(무처치, 단일처치, 이중처치)를 가정하였다.

혼합 설계 분산분석 결과 mixed design ANOVA(그림 2-4), 이전 행동의 결과(보상 여부)의 주효과가 유의한 것으로 나타났고($F_{(1,35)} = 32.20, p < 0.001$), 이는 과제 수행에 모형 없이 model-free 행동을 선택하는 습관 체계가 영향을 미쳤음을 시사한다. 또한 행동의 결과(보상 여부)와 과제 구조(단계 전환 형태) 간의 상호작용이 유의미하게 나타났고($F_{(1,35)} = 11.73, p = 0.002$) 이는 과제 수행에 모형 부재 model-free 행동 체계뿐 아니라 모형 기반의 model-based 행동 체계 역시 영향을 미쳤음을 시사한다.

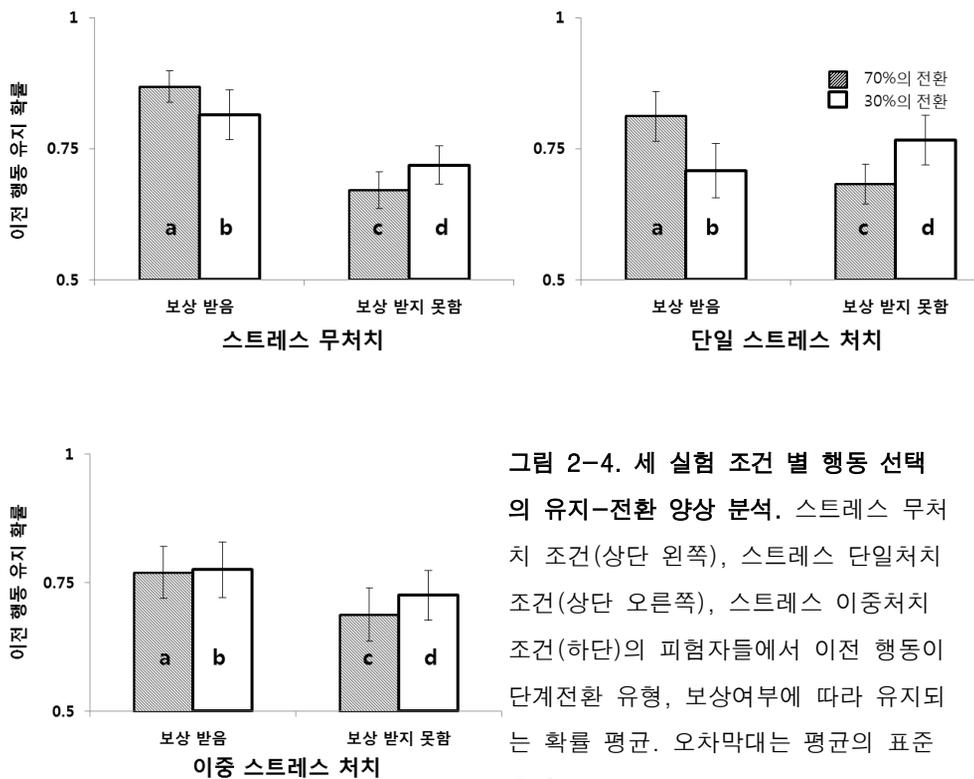


그림 2-4. 세 실험 조건 별 행동 선택의 유지-전환 양상 분석. 스트레스 무처치 조건(상단 왼쪽), 스트레스 단일처치 조건(상단 오른쪽), 스트레스 이중처치 조건(하단)의 피험자들에서 이전 행동이 단계전환 유형, 보상여부에 따라 유지되는 확률 평균. 오차막대는 평균의 표준 오차.

이 두 요인 간 상호작용 효과와 스트레스 간의 상호작용은 유의하게 나타나지 않았으며 ($F_{(2,35)} = 2.05, p = 0.144$), 이전 행동에 대한 보상 여부와 스트레스 간 상호작용 효과는 유의하게 나타났는데 ($F_{(2,35)} = 5.30, p = 0.010$), 이는 스트레스 무처리 조건에서 보상받은 행동을 유지한 비율이 다른 조건에 비해 높은 것에 기인하는 것으로 보인다.

2) 강화학습에 대한 계산 모형 적용

다음으로는 피험자들의 의사결정 수행 자료에 강화학습 모형 중 RL model 2 를 적용하여, 피험자 별 의사결정 특성에 대한 모수들을 추정했다. RL model 2 의 모수 최대우도추정치 maximum likelihood estimate 들의 실험 조건 별 사분위 값들이 표 2-2, 모수추정치들과 행동 측정치들간의 상관관계가 표 2-3 에 제시되어 있다.

표 2-2. RL model 2 모수 최대우도 추정치들의 조건 별 사분위 값들

	γ	Δ_{+}^{mf}	Δ_{+}^{mb}	Δ_{-}^{mf}	Δ_{-}^{mb}	$\Delta_{+}^{mb} - \Delta_{-}^{mb}$	$\Delta_{+}^{mf} - \Delta_{-}^{mf}$
무처치							
25th	0.45	0.31	-0.14	-0.17	-0.19	-0.15	0.66
Median	0.54	1.76	-0.01	0.06	-0.13	0.01	1.04
75th	0.93	2.32	0.29	0.35	0.05	0.22	2.17
단일처치							
25th	0.48	0.38	-0.06	-0.05	-0.69	0.11	-0.07
Median	0.68	0.69	0.12	0.24	-0.40	0.61	0.36
75th	0.90	1.16	0.92	0.97	-0.10	1.34	0.86
이중처치							
25th	0.66	0.01	-0.13	0.03	-0.23	-0.02	-0.06
Median	0.82	0.45	0.01	0.28	-0.16	0.09	0.43
75th	0.89	1.02	0.16	0.42	0.02	0.30	0.70

표 2-3. 모수의 최대우도 추정치들과 행동 측정치들 간의 상관계수

	p_correct	outcomes	subblock_2	P_mf	P_mb
Γ	-.32 (.047)	.00 (.983)	-.34 (.039)	-.14 (.418)	-.26 (.112)
Δ_{+}^{mf}	.27 (.098)	.34 (.038)	.32 (.049)	.46 (.003)	-.09 (.584)
Δ_{+}^{mb}	.36 (.026)	.10 (.553)	.38 (.020)	-.40 (.012)	.70 (<.001)
Δ_{-}^{mf}	.19 (.262)	.39 (.017)	.12 (.476)	-.44 (.006)	.23 (.156)
Δ_{-}^{mb}	-.39 (.017)	-.65 (<.001)	-.31 (.056)	.23 (.161)	-.50 (.001)
$\Delta_{+}^{mb} - \Delta_{-}^{mb}$.44 (.005)	.36 (.026)	.447 (.005)	-.41 (.011)	.75 (<.001)
$\Delta_{+}^{mf} - \Delta_{-}^{mf}$.17 (.316)	.123 (.461)	.25 (.125)	.70 (<.001)	-.22 (.183)

P_correct: 구간 2 ~ 5 에서의 유리한 옵션 선택 비율; outcomes: 구간 2 ~ 5 에서 얻은 점수 총합; 구간 3 ~ 5 의 각 구간 내 후반부에서의 유리한 옵션 선택 비율(반전 학습 정도); P_mf = (a + b) - (c + d); P_mb = (a + d) - (b + c) (그림 2-4 참조)

RL model 2 의 모수 최대우도추정치들 maximum likelihood estimators 의 세 조건 간 차이를 일원분산분석 ANOVA 한 결과, 보상 받지 못한 행동에 대해서 i) 모형 기반 model-based 방식으로 행동을 전환하는 양상인 \mathcal{A}_{-}^{mb} ($F_{(2,35)} = 3.68, p = 0.035$)과 ii) 모형 기반 model-based 행동 양상을 반영하는 \mathcal{A}_{+}^{mb} 와 \mathcal{A}_{-}^{mb} 간 차이($\mathcal{A}_{+}^{mb} - \mathcal{A}_{-}^{mb}; F_{(2,35)} = 3.87, p = 0.030$) 및 iii) 모형 부재 model-free 행동 양상을 반영하는 \mathcal{A}_{+}^{mf} 와 \mathcal{A}_{-}^{mf} 간 차이($\mathcal{A}_{+}^{mf} - \mathcal{A}_{-}^{mf}; F_{(2,35)} = 3.91, p = 0.029$)의 추정치들에서 조건 간 차이가 유의하게 나타났다.

사후분석으로 Turkey-HSD 검증을 통해 다중비교한 결과, 스트레스 무처치 조건은 단일처치 조건에 비해 \mathcal{A}_{-}^{mb} 추정치가 높게 나타났으며($p = 0.034$), $\mathcal{A}_{+}^{mb} - \mathcal{A}_{-}^{mb}$ 값이 유의하게 낮았고($p = 0.048$), $\mathcal{A}_{+}^{mf} - \mathcal{A}_{-}^{mf}$ 값은 높았다($p = .041$). 한편 스트레스 이중처치 조건의 피험자들은 무처치 조건에 비해 $\mathcal{A}_{+}^{mb} - \mathcal{A}_{-}^{mb}$ 값이 낮게 나타났으나($p = .065$) 통계적으로 유의한 수준엔 못 미쳤고, 다른 모수 추정치들에서의 조건 간 차이는 나타나지 않았다. \mathcal{A}_{+}^{mf} 와 \mathcal{A}_{-}^{mf} 추정치 간 차이는 유의하지 않았다($t_{23} = -0.23, p = 0.822$).

이어서, 양방향 t -검증을 통해 실험 조건들 간의 차이를 쌍비교해 보았다. 그 결과, 스트레스 무처치 집단은 단일처치 집단에 비해 \mathcal{A}_{-}^{mb} 추정치가 높게 나타났으며($t_{24} = 2.38, p = 0.027$), $\mathcal{A}_{+}^{mb} - \mathcal{A}_{-}^{mb}$ 값이 유의하게 작게 나타났고($t_{24} = -2.19, p = 0.039$), $\mathcal{A}_{+}^{mf} - \mathcal{A}_{-}^{mf}$ 는 큰 것으로 나타났다($t_{24} = 2.65, p = 0.015$). 반면 스트레스 이중처치

집단의 피험자들은 단일처치 집단에 비해 $\Delta_{+}^{mb} - \Delta_{-}^{mb}$ 값이 유의하게 작게 나타났고($t_{23} = 2.30, p = 0.031$), Δ_{+}^{mf} 와 Δ_{-}^{mf} 추정치 간 차이는 유의하지 않았다($t_{23} = -0.23, p = 0.822$). 한편 옵션의 이전 기대치에 대한 가중치(망각률)인 γ 추정치에서는 집단 간 차이가 유의하지 않았다.

이러한 결과는 스트레스 단일처치를 받은 피험자들은 무처치 피험자들에 비해 모형 기반 model-based 행동 양상이 증가하고 모형 부재 model-free 행동 양상이 감소된 반면, 스트레스의 이중처치를 받은 피험자들은 단일처치 받은 피험자들에 비해 모형 기반의 model-based 행동 양상이 저하되었음을 뜻한다. 즉 스트레스는 처치 수준에 따라 모형 기반 model-based 행동에 양방향적인 bidirectional 영향을 행사한 것으로 보인다(그림 2-5).

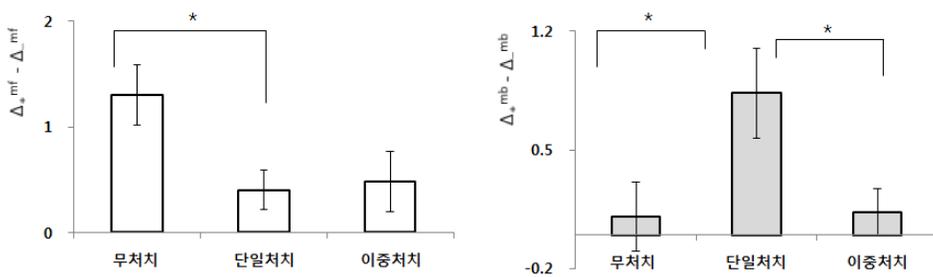
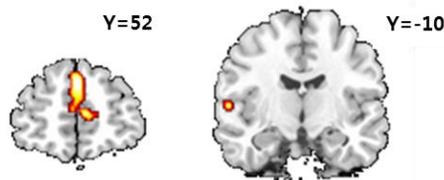


그림 2-5. 모형 기반 행동 선택에 대한 스트레스의 처치 수준 별 효과. * $p < .05$

3) 의사결정 관련 신경활성화에 대한 스트레스의 양방향적 영향

다음으로는 모형 기반 model-based 행동 선택에 대한 스트레스의 양방향적인 영향을 뇌신경활동 수준에서 탐색해보았다. 우선, 스트레스 단일처치 조건의 피험자들에서 스트레스 무처치 조건과 신경활성화가 다르게 나타나는 영역을 탐색(양방향 t -검증)한 결과, 스트레스 무처치 조건이 스트레스 단일처치 조건에 비해 유의미하게 더 활성화된 영역은 없었고, 스트레스 단일처치 조건이 스트레스 무처치 조건에 비해 더 활성화된 영역은 양측 내측 전전두엽 medial prefrontal cortex (Brodmann Area, 이하 BA, 10)과 좌측 상측두엽 superior temporal gyrus (BA 22)으로 나타났다($p_{\text{uncorrected}} < .001$; $k > 50$) (그림 2-6).



영역	BA	k	MNI			Z 값
			x	Y	z	
내측 전전두엽	10	920	0	52	30	4.22
상측 측두엽	22	100	-60	-10	6	3.81

그림 2-6. 첫 번째 의사결정 시 스트레스 단일처치로 인해 활성화가 증가된 내측 전전두

엽 medial prefrontal cortex 과 상측두엽 superior temporal cortex 영역 (역치: $p_{uncorrected} < .001$, 영역 크기 > 50; BA: Brodmann Area)

앞선 행동 자료 분석에서 모형 기반 model-based 행동이 스트레스 단일처치 시 증가했다가 이중처치 시 다시 감소한 것으로 나타난 결과를 고려할 때, 위 두 관심 영역에서의 스트레스 단일처치로 인한 신경활성화 증가가 이중처치로 인해 감소할 가능성이 예상된다. 따라서 위 두 영역을 첫 번째 의사결정 시 스트레스 단일 처치로 인해 신경활성화가 증진된 영역으로 간주하고 관심 영역 region of interest 으로 지정하여, 스트레스 단일처치 조건과 이중처치 조건 간의 신경활성화 차이를 확인해보았다. 스트레스 단일처치 조건과 이중처치 조건 간 신경활성화 차이를 두 관심 영역 region of interest 에서 양방향 t -검증한 결과, 스트레스 이중처치 조건은 단일처치 조건에 비해 두 관심 영역 내의 내측 전전두엽 medial prefrontal cortex 과 좌측 상측두엽 superior temporal cortex 영역에서 신경활성화가 저조한 것으로 나타났다($p_{corrected} < .05$, family-wise error correction)(그림 2-7A, B).

상기 결과는, 의사결정자가 자신이 처한 환경을 파악하여 이용하는 모형기반 model-based 행동 선택 시 내측 전전두엽 영역과 좌측 측두엽의 신경활성화가 경미한 스트레스로는 증진되나, 부가적인 스트레스가 가해면 저하되는 것을 보여준다. 이는 내측 전전두엽 영역과 좌측 측두엽의 신경활성화에 있어서 스트레스의 양방향적 영향을 시사한다.

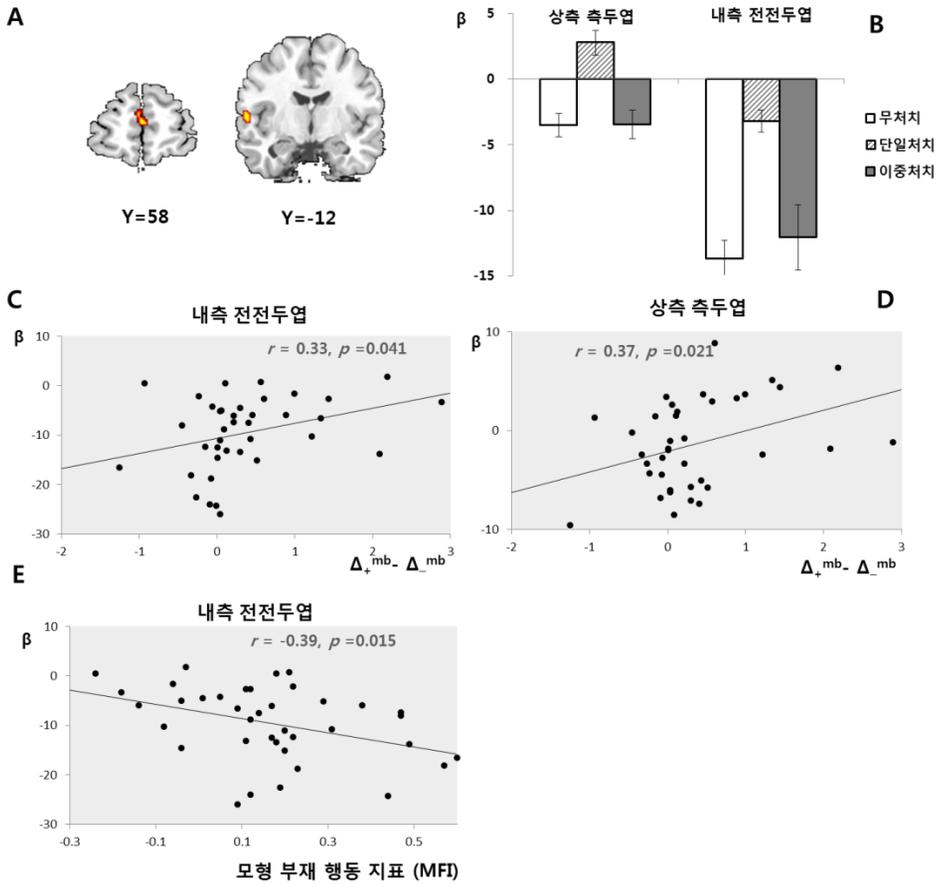


그림 2-7. 의사결정 관련 신경활성화에 대한 스트레스의 양방향 효과. A. 스트레스 단일처치로 신경활성화가 증가하고 이중처치로 저하하는 좌측 상측두(superior temporal) 영역(중심좌표: MNI [-58, -12, 6])과 내측 전전두(medial prefrontal) 영역(MNI [0, 58, 8]) ($P_{FWEcorrected} < .05; k \geq 50$). B. A의 두 영역에서 추출한, 첫 번째 의사결정 사건의 GLM 회귀계수 추정치의 조건 별 평균(오차막대는 표준오차). C, D. 내측 전전두엽과 좌측 상측두 영역의 의사결정 관련 신호의 회귀계수 추정치와 RL model 2의 모형기반 행동 관련 모수추정치 간 상관관계. E. 모형부재 행동 지표값과 내측 전전두엽의 의사결정 관련 회귀계수 추정치 간 부적 상관관계.

다음으로는 의사결정 관련 뇌 활성화에 대한 일반선형모형을 통해 상기 두 영역에서 추정된 회귀계수치(첫 번째 의사결정 사건의 회귀계수 추정치)를 추출하여, RL model 2에서 추정된 모형 기반 model-based 행동의 모수 추정치($\beta_+^{mb} - \beta_-^{mb}$)와의 상관관계를 분석했다. 그 결과, 내측 전전두엽 medial prefrontal cortex 영역의 회귀계수 모수치와 모형 기반 model-based 행동을 반영하는 모형 모수 추정치들로 계산한 ' $\beta_+^{mb} - \beta_-^{mb}$ '와의 상관관계가 유의했으며($r = .33, p = .041$) (그림 2-7C), 좌측 상측두엽 superior temporal cortex 영역의 회귀계수 모수치와 ' $\beta_+^{mb} - \beta_-^{mb}$ ' 간의 상관관계 역시 유의한 것으로 나타났다($r = .37, p = .021$) (그림 2-7D). 또한 내측 전전두엽 medial prefrontal cortex 의 신경활성화는 과제 수행 결과, 보상 여부에 따라 이전 행동을 유지하는 비율(MFI = (a + b) - (c + d), 그림 2-4)과 부적 상관관계로 나타났다($r = -.39, p = .015$) (그림 2-7E). 이러한 결과는, 의사결정 시 스트레스가 처치 정도에 따라 양방향적 영향을 행사하는 것에 대한 신경활성적 기반으로 보이며, 특히 내측 전전두엽 medial prefrontal cortex 영역은 모형 기반 model-based 행동 경향뿐 아니라 모형 부재 model-free 행동 경향을 조절하는 역할을 할 가능성을 시사한다.

한편, 의사결정 과제 수행 중, 결과 제시 사건 event 관련 신경활성화 양상에서는 세 조건 간 차이가 유의하지 않았다.

4) 강화학습 모형 기반의 뇌 영상 분석

강화학습 모형 RL model 2 를 기반으로 매 시행의 선택 행동의 기대치 chosen value 와 결과 예측 오류 prediction error (선택 행동의 기대치와 실제 결과 가치 간의 차이)를 구하여, 뇌의 어느 영역에서의 신경활동 변화와 관련성을 갖는지 탐색하고 그에 대한 조건 간 차이를 검증했다. 피험자 개인마다 과제 수행 시 매 시행의 선택 행동의 기대치 chosen value 와 결과 예측 오류 prediction error 를 계산하기 위해, 각 조건 내에서 모수 추정치의 고정 효과 fixed effect 를 가정하여, 조건 별로 RL model 2 의 모수들을 추정했다. 조건 별 모수 추정 결과는 표준편차($\text{asymptotic covariance estimator}$ 로 계산; Hessian 역행렬의 대각선 요소들의 제곱근에 해당)와 함께 표 2-4A 에 제시했다. 이 결과, 즉 조건 간 차이와 조건 내 고정효과를 가정한 모형의 추정 결과(부적 로그 우도 값=3481.37, BIC=7095.91)는, 전체 조건들 내의 고정효과를 가정한 모형(부적 로그 우도 값=3534.99, BIC=7114.38, 표 2-4B)에 비해, 행동 자료에 대한 모형 적합도가 좋은 것으로 나타났다.

표 2-4. 뇌 영상 분석을 위한 모수 추정 결과

A. 각 실험 조건 별로 추정된 모수치 ± 표준편차						
	-log likelihood	γ	Δ_{+}^{mf}	Δ_{+}^{mb}	Δ_{-}^{mf}	Δ_{-}^{mb}
무처치	2471	0.56±0.03	1.28±0.09	0.08±0.08	0.32±0.06	-0.07±0.05
단일처치	2482	0.71±0.02	0.74±0.07	0.18±0.05	0.38±0.05	-0.15±0.05
이중처치	2224	0.80±0.02	0.68±0.06	-0.11±0.05	0.31±0.05	-0.09±0.05
B. 전체 피험자에서 추정된 모수치 ± 표준편차						
	-log likelihood	γ	Δ_{+}^{mf}	Δ_{+}^{mb}	Δ_{-}^{mf}	Δ_{-}^{mb}
전체	7177	0.69±0.01	0.89±0.04	0.05±0.03	0.34±0.03	-0.09±0.03

선택 행동의 기대치 관련 뇌 활성화에 대한 스트레스 효과. 강화학습 모형 RL model 2에서, 세 실험 조건 별로 추정된 모수치들을 이용하여 각 피험자들의 과제 수행 시 매 시행마다의 trial-by-trial 선택 행동의 기대치 $\text{trial-by-trial expected values of an action selected}$ 와 예측오류 prediction error 를 계산했고, 이 수치들을 각각 BOLD 신호에 대한 첫 번째 의사결정 사건 event 과 결과 제시 사건 event 의 조절인자 $\text{parametric modulator}$ 로 가정하여, 새로운 회귀변수 regressor 로 일반선형모형 GLM 에 투입하였다.

세 실험 조건(스트레스 무처치, 단일처치, 이중처치) 각각에서 한 집단 t -검증을 실시하며 선택 행동의 기대치 chosen value 관련 뇌 활성화 양상을 확인한 결과, 스트레스 무처치 조건에서는 우측 중심전회 precentral gyrus 와 우측 해마 hippocampus 의 신경활성화가 선택 행동의 기대치 chosen value 와 정적 상관인 것으로 나타났으며(그림 2-8, A), 상측 두정엽 $\text{superior parietal cortex}$ 의 신경활성화는 부적인 상관이 유의했다(그림 2-8, B). 반면

스트레스 단일처치 및 이중처치 조건에서는 정적 상관관계가 유의한 영역이 나타나지 않았으며, 두 조건 모두에서 췌기앞소엽_{precuneus}의 신경활성화와의 부적 상관관계가 유의했다(그림 2-8, C, D).

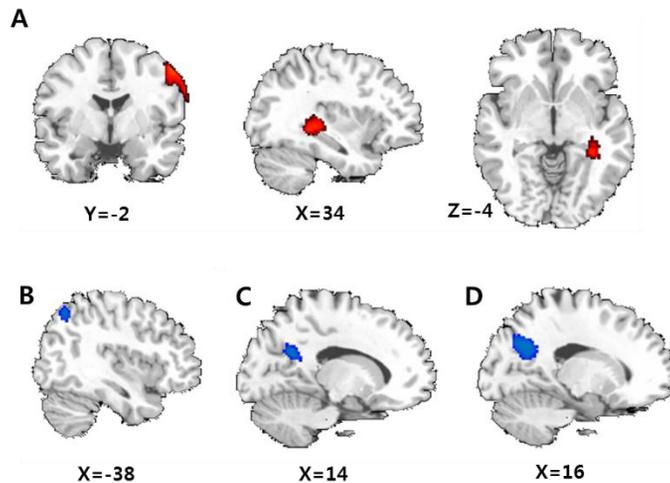


그림 2-8. 실험 조건 별 선택 행동의 기대치 관련 BOLD 신호가 유의하게 나타난 영역.
 A. 스트레스 무처치 조건에서 정적으로 유의하게 나타난 중심전회_{precentral gyrus}(중심좌표 MNI [54, -2, 50], $z=4.51$), 해마_{hippocampus}(중심좌표 MNI [34, -38, -4], $z=4.16$); B. 스트레스 무처치 조건에서 부적으로 유의하게 나타난 상측 두정엽_{superior parietal cortex}(중심좌표 MNI [-38, -68, 52], $z=3.73$); C. 스트레스 단일처치 조건에서 부적으로 유의하게 나타난 췌기앞소엽_{precuneus}(중심좌표 MNI [14, -56, 24], $z=4.14$); D. 스트레스 이중처치 조건에서 부적으로 유의하게 나타난 췌기앞소엽(중심좌표 MNI [16, -66, 28], $z=4.21$); 역치: $P_{uncorrected} < .001$, 군집 크기 > 100 , $z\text{-value} > 3.5$.

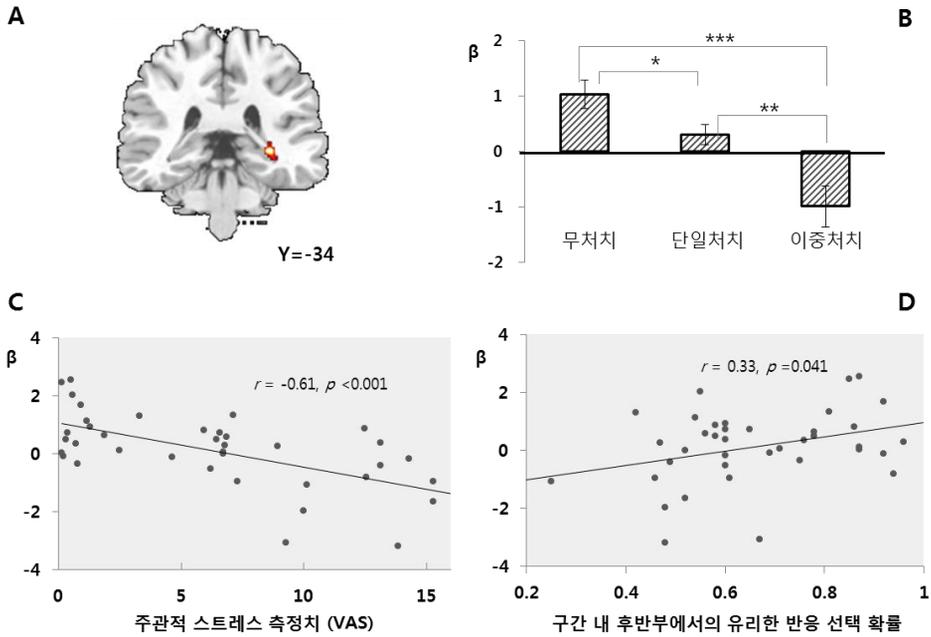


그림 2-9. 선택 행동의 기대치 관련 신경활성화가 나타난 영역 중 실험 조건 간 차이가 유의한 영역. A. 우측 해마_{hippocampus} (MNI 좌표 [36, -30, 2], $p < .001$, 군집크기 $k = 23$), B. 우측 해마_{hippocampus}의 회귀계수 추정치들의 실험 조건 간 평균(오차 막대는 표준 오차), * $p = .053$, ** $p = .007$, *** $p < .001$. C. 우측 해마_{hippocampus}에서 추출한 회귀계수 추정치와 주관적 스트레스 측정치 간 부적 상관관계. D. 우측 해마_{hippocampus}에서 추출한 회귀계수 추정치와 구간 내 후반부에서 유리한 반응을 선택할 확률 간의 상관관계.

다음으로는 세 실험 조건 간 차이를 탐색하기 위해 ANOVA 분석을 실시했다. 그 결과, 우측 해마_{hippocampus}의 신경활성화에서 세 조건 간 차이가 유의미한 것으로 나타났다(그림 2-9, A). 이 영역에서 피험자별 선택 행동 가치_{chosen value}의 GLM 회귀계수 모수추정치를 추출하여

실험 조건 간 차이를 검증했다. 양방향 t 검증 결과, BOLD 신호에 대한 선택 가치의 회귀계수 추정치는 스트레스 처치에 따라 감소했으며(그림 2-9, B), 주관적 스트레스 측정치(VAS 점수)가 높을수록 저조하게 나타났다(그림 2-9, C). 또한 선택 가치 관련 회귀계수 추정치들은 반전 학습 구간(구간 3 ~ 6)에서 구간 내 후반부의 유리한 옵션 선택 확률과는 정적 상관관계인 것으로 나타났다(그림 2-9, D). 이는 스트레스가 의사결정 과정에서 우측 해마의 선택 행동의 기대치 반영을 방해하고 이로 인해 반전 학습에 부정적인 영향을 미칠 가능성을 시사한다.

예측 오류 관련 뇌 활성화에서의 스트레스 효과. 피험자 별로 각 시행마다 예측 오류(prediction error)를 계산하여 BOLD 신호와의 상관관계가 유의한 영역을 탐색했다(그림 2-10, 표 2-5). 그 결과, 스트레스 무처치 조건에서는 하측 후두엽(inferior occipital cortex), 전측 대상회(anterior cingulate cortex), 중측 대상회(middle cingulate cortex), 후측 대상회(posterior cingulate cortex), 조거구(calcarine sulcus), 각회(angular gyrus), 미상핵(caudate nucleus), 하측 두정엽(inferior parietal cortex)에서 정적 상관관계가 유의하게 나타났다(전뇌 수준에서 family-wise error correction 실시 결과 $p_{\text{corrected}} < .05$ 에서 유의). 스트레스 단일처치 조건에서는 전뇌 교정 후 $p < .05$ 수준에서 유의한 영역은 없었으나, 미상핵(caudate nucleus), 전측 대상회(anterior cingulate cortex), 후측 대상회(posterior cingulate cortex), 피각(putamen), 상측 두정엽(superior parietal cortex), 하측 후두엽(inferior occipital cortex)에서 $p_{\text{uncorrected}} < .001$ 수준의 유의한 정적

상관관계가 확인되었고, 스트레스 이중 처치 조건에서는 후두엽occipital cortex의 정적 신경활성화가 관찰되었다. 세 조건 모두에서 예측 오류prediction error 관련 신경활성화가 부적으로 유의한 영역은 관찰되지 않았으며, 조건 간 차이를 분석한 결과에서 세 조건 간 차이가 유의한 영역 역시 없었다.

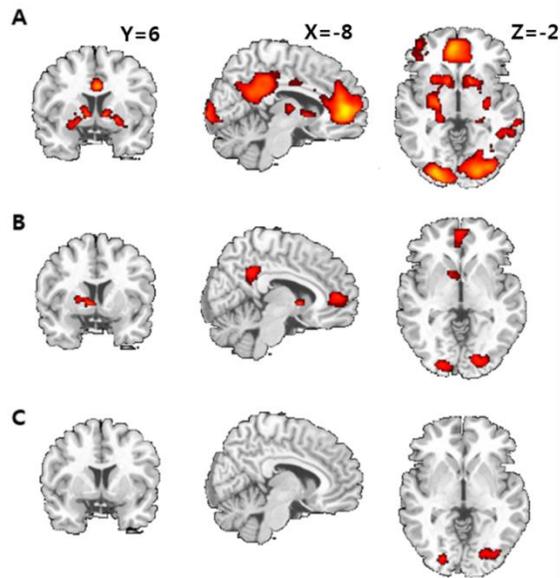


그림 2-10. 실험 조건 별 예측 오류 관련 BOLD 신호가 유의하게 나타난 영역. A. 스트레스 무처치 조건: 하측 후두엽, 전측대상회, 상측 전두엽, 조거구calcarine, 중대상회middle cingulum, 후대상회posterior cingulum, 피각putamen, 미상핵caudate nucleus, 각회angular gyrus; B. 스트레스 단일처치 조건: 피각, 미상핵, 각회, 전측대상회, 중대상회, 내측 안와전두엽orbital medial frontal gyrus; C. 스트레스 이중처치 조건: 하측 후두엽; 역치: $p_{uncorrected} < .001$, 군집 크기 > 100 , z -value > 3.5 .

표 2-5. 실험 조건 별 예측 오류 관련 신경활성화가 나타난 영역들: 한 집단 t-검증 결과

영역	BA	k	MNI			Z 값
			x	y	z	
스트레스 무처치 조건 ($p_{corrected} < .05$, FWE; $k > 15$; $z > 4.5$)						
하측 후두엽 inferior occipital cortex	17	390	-26	-90	-6	5.88
전측 대상회 anterior cingulate cortex	32	933	-8	44	4	5.77
조거구 calcarine		357	20	-90	0	5.24
중측 대상회 middle cingulate cortex	24	105	2	2	32	5.23
후측 대상회 posterior cingulate cortex		280	2	-44	28	5.06
미상핵 caudate nucleus		17	10	8	-2	5.01
하측 두정엽 inferior parietal cortex	40	37	-60	-44	46	4.93
각회 angular gyrus	39	24	-50	-68	38	4.75
스트레스 단일처치 조건 ($p_{uncorrected} < .001$; $k > 20$)						
미상핵 caudate nucleus		80	10	8	-2	4.49
미상핵 caudate nucleus		121	-8	6	-2	4.27
상측 두정엽 superior parietal cortex	7	550	-36	-72	50	4.26
전측 대상회 anterior cingulate cortex	32	458	-6	48	6	3.96
중측 대상회 middle cingulate cortex	31	314	-8	-46	34	3.85
하측 후두엽 inferior occipital cortex		433	-22	-92	-4	3.72
피각 putamen		27	24	8	-8	3.62
스트레스 이중처치 조건 ($p_{uncorrected} < .001$; $k > 20$)						
중측 후두엽 middle occipital cortex		172	24	-80	2	3.80
하측 후두엽 inferior occipital cortex	18	224	-34	-86	-12	3.71

BA: 브로드만 영역; k: 군집 크기

3. 논의 2

연구 2에서는 스트레스가 의사결정에 미치는 인지행동적 영향을 신경활동 수준에서 탐색하는 동시에, 스트레스 처치 정도에 따른 의사결정에 대한 영향을 탐색했다. 연구 결과, 스트레스는 처치 수준에 따라 의사결정에 양방향의bidirectional 영향을 미치는 것으로 나타났으며, 그 신경활동적 기전은 스트레스의 정도에 따른 내측 전전두엽medial prefrontal cortex과 상측 측두엽superior temporal cortex 신경활성화의 양방향적bidirectional 변화인 것으로 나타났다. 또한 스트레스 처치는 우측 해마hippocampus에서 나타나는 선택 행동의 기대치chosen value 관련 BOLD 신호를 저하시켰으며 이는 행동적으로는 반전 학습 저하로 나타나, 스트레스가 의사결정을 위한 강화학습 과정에 신경활동 및 인지행동적 수준에서 직접적인 영향을 미치는 것이 확인되었다.

1) 스트레스의 양방향적 조절 효과와 신경활동적 기전

스트레스가 처치 정도에 따라 기억 기능에 양방향적 영향을 미친다는 것은 선행 실험 연구들에서 밝혀진 바 있다(Gold et al., 1975; Diamond et al., 1992; Sauro et al., 2003; Roozendaal et al., 2009; Salehi et al., 2010). 즉 스트레스는 경미한 정도일 때 기억 기능

과제에서 수행 증진을 초래할 수 있으나, 정도가 심한 경우에는 수행 저하를 일으킨다. 그러나 스트레스가 행동 선택에 미치는 영향은 목표지향적인 행동을 감소시키고 습관적인 행동을 증진시키는 단조적인 방식인 것으로 보고되어 왔을 뿐, 스트레스 처치 정도에 따라 그 영향이 달라지는지, 달라진다면 어떻게 달라지는지 등에 대해서는 아직 명확히 밝혀지지 않았다.

본 연구 결과, 스트레스는 처치 정도에 따라 의사결정에 양방향적인bidirectional 영향을 행사한 것으로 나타났다. 즉, 스트레스 단일처치 조건의 피험자들은 무처치 조건에 비해 오히려 모형 기반의 model-based 목표지향적 행동이 증가했고 모형 부재 model-free 의 행동 경향이 감소되었으나, 스트레스 수준이 더 높은 스트레스 이중처치 조건은 단일처치 조건에 비해 모형 기반model-based 행동이 저조하게 나타났다. 이러한 양방향적 스트레스 영향은 뇌신경 활동 수준에서도 확인되었다. 즉 스트레스 단일처치는 의사결정 시 내측 전전두엽 medial prefrontal cortex과 좌측 상측두엽superior temporal cortex의 신경활성화를 증진시켰으나, 스트레스 이중처치는 이 영역들의 신경활성화를 감소시킨 것으로 나타났다.

내측 전전두엽medial prefrontal cortex과 좌측 상측두엽superior temporal cortex의 신경활성화는 강화학습 모형reinforcement learning model에서 계산된 모형 기반model-based 행동 모수치의 크기와 신경활성화 정도가 비례하는 것으로 나타나, 목표지향적 행동을 조절하는 주요 영역인 것으로 보인다.

상측 측두엽(superior temporal cortex)은 의미 처리, 청각 자극 처리에 중요역할을 하는 것으로 알려져 왔으나, 최근에는 목표지향적인 움직임 관찰하거나 목표지향적으로 움직이는 데에 관여하는 것으로 밝혀졌다(O'Doherty et al., 2004). 또한 인간을 대상으로 한 뇌 영상 연구에서는 컴퓨터와의 경쟁적 가위바위보 게임 중 이전 행동의 결과가 '승(win)'일 때 상측 측두엽(superior temporal cortex)이 활성화되며, 특히 뇌도(insula)와의 연결 부위는 이전 수행의 행동을 유지할 때보다 다른 행동으로 바꿀 때 활성화되는 것을 확인했다(Paulus et al., 2005). 이러한 결과를 종합해볼 때, 상측 측두엽은 이전 행동과 그 결과를 통합하여 목표지향적으로 행동을 선택해나가는 데에 중요한 역할을 담당하는 것으로 보인다. 본 연구에서는 상측 측두엽의 신경활성화가 모형 기반(model-based) 행동 경향과 정적 상관을 이루는 것을 확인한 동시에 이러한 모형 기반(model-based) 행동 관련 신경활성화가 스트레스의 처치 수준에 따라 양방향적으로 변화하는 것을 확인함으로써, 이 영역이 스트레스가 의사결정에 미치는 영향에 대한 기저 신경학적 현상을 설명하는 주요 영역임을 제시한다.

또한 내측 전전두엽(medial prefrontal cortex)은 의사결정에 중요한 역할을 하는 뇌 영역이며, 그 중 복내측 영역(ventromedial prefrontal cortex)은 선택 행동의 가치(value of action selected) 및 예측 오류(prediction error)를 반영하여, 행위자의 내적 동기 수준에 따라 주어진 상황 내에서 행동을 선택하는 것에 관여한다(Dickerson et al., 2011). 선행 실험 연구들에서 복내측

전전두엽(ventromedial prefrontal cortex)이 행동 결과 자체의 가치를 반영할 뿐 아니라, 선택한 행동의 기대치(chosen value)를 입력하는 영역임이 밝혀졌다(Schönberg et al., 2007; Gläscher et al., 2009). 특히 최근 한 뇌 영상 연구에서는 인간 뇌에서 복내측 전전두엽(ventromedial prefrontal cortex)의 신경활성화가 습관행동 조형의 주요 영역인 후측 피각(posterior putamen)의 신경활성화와 부적 상관관계를 맺는 것을 확인하여(Lee et al., 2014), 복내측 전전두엽(ventromedial prefrontal cortex)이 의사결정 중 습관적 행동을 통제함으로써 목표지향적 행동 시스템과 습관 행동 시스템 간 균형에 조절자(arbitrator) 역할을 할 수 있음을 제시했다. 본 연구 결과에서는 상기 연구의 복내측 전전두엽(ventromedial prefrontal cortex) 영역과 (MNI 좌표 상 위치가) 일치하는 영역에서 선택 행동의 기대치(chosen value) 관련 신경활성화 정도가, 모형 기반(model-based) 행동 모수치와 정적 상관관계인 동시에, 모형 부재(model-free) 행동 지표와 부적 상관관계로 나타났다. 이는 스트레스가 복내측 전전두엽(ventromedial prefrontal cortex) 영역의 신경활성화를 변화시킴으로써 의사결정의 두 시스템 간 균형 및 조절에 관여할 가능성을 시사한다.

2) 의사결정에 있어서 해마의 역할과, 그에 대한 스트레스의 영향

본 연구에서는 강화학습 모형에서 추정된 모수들을 이용하여,

피험자들이 과제 수행을 하는 과정에서 매 시행 선택하는 옵션에 대한 기대치를 계산하여 의사결정으로 인한 신경활성화와의 상관관계를 탐색하였다. 그 결과, 선택 행동의 기대치를 반영하는 것으로 나타난 뇌 영역 중 우측 해마hippocampus의 신경활성화는, 스트레스로 인해 활성화 정도가 달라지는 것으로 나타났다. 즉 우측 해마에서의 선택 행동의 기대치chosen value 관련 BOLD 신호는 스트레스 처치에 따라 감소했으며(무처치 > 단일처치 > 이중처치), 피험자들의 주관적인 스트레스 측정치VAS score와는 비례했다. 이러한 결과는 의사결정 과정에 있어서 해마hippocampus의 역할에 스트레스가 부정적인 영향을 미침을 시사한다.

해마hippocampus는 행동 결과의 가치에 기반하는 강화학습 및 의사결정에서 단순히 행동의 보상 여부를 기억하는 것 이상의 중요한 역할을 한다(Dickerson et al., 2011; Foerde and Shohamy, 2011). 확률적 의사결정 과제와 같이 인지적 노력이 요구되는 행동 선택을 수행할 때에는, 행동과 그 결과에 대한 경험이 일화적 기억episodic memory으로 축적되는 것이 중요하다. 매 시행의 행동-결과의 경험이 이전에 얻은 정보들과 통합되어서 축적되어 다음 행동 선택에 반영되어야 하므로, 일화적 기억 기능episodic memory이 강화학습reinforcement learning에 중요한 역할을 할 수 있다. 이와 관련하여, 선행 연구에서는 동물 해마hippocampus의 CA1에 강화학습 모형reinforcement learning model에서 계산된 선택 행동의 기대치chosen value, 행동의 결과outcome, 선택 행동action

selected 자체에 대한 신호가 모두 반영됨이 확인되었다(Lee et al., 2012). 인간 연구에서도 해마hippocampus는 강화학습reinforcement learning에서 선택 행동의 가치chosen value를 반영함이 확인되었으며(Adcock et al., 2006; Wimmer et al., 2012), 이러한 연구 결과들은 강화학습reinforcement learning 과정 중 경험하는 정보가 해마hippocampus에서 통합되어 행동 선택에 활용됨을 뜻한다. 코티솔 수용체가 밀집되어 있는 해마hippocampus는 스트레스 반응을 조절하는 주요 뇌 영역인 동시에, 스트레스의 영향에 구조적으로나 기능적으로 가장 민감한 영역이다(McEwen and Sapolsky, 1995; Lupien and Lepage, 2001; Kim and Diamond, 2002). 본 연구에서는 스트레스 처치로 인해 해마hippocampus에서 선택 행동의 기대치chosen value 관련 BOLD 신호가 저하되는 것을 확인함으로써, 해마에 있어서 스트레스의 영향이 기억 기능 저하에 국한되지 않고, 강화 학습 및 의사결정의 과정에도 부정적인 영향을 미칠 수 있음을 밝혔다.

한편 본 연구 결과에서 스트레스와 관련된 것으로 나타난 우측 해마의 의사결정 관련 신경활성화는, 반전 학습reversal learning 수행 정도와 정적 상관관계로 나타났다. 즉, 스트레스 처치의 영향을 받는 것으로 보이는 우측 해마hippocampus의 선택 행동의 기대치chosen value 관련 BOLD 신호는 스트레스 처치 조건의 피험자들의 반전학습reversal learning 수행이 저조한 것과 일치했다. 반전 학습reversal learning은 과제 수행 중 구간block마다 단서 없이 행동-결과 간 수반성contingency이 반전되는 실험

패러다임이다. 반전 학습 reversal learning 과제에서 행동과 결과 간 관계가 100%의 결정적 deterministic 관계라면, 행위자들은 수반성 contingency 이 반전되는 양상을 비교적 초반에 습득할 수 있을 것이다. 그러나 확률적 반전 학습 과제 probabilistic reversal learning task 에서는 행동을 선택하고 그 결과를 확인하는 경험이 지속적으로 축적되면서 과제 중반 이후에 반전 학습 reversal learning 양상을 습득하게 된다. 즉 행동과 보상 간 관계가 확률적인 과제에서는 과제 중 ‘행동-결과 간 수반성 contingency 이 반전된다’는 정보는 과제 초반에 습득하여 수행 중 이용할 수 있는 정보가 아니며, 이때 행위자가 파악해야 하는 것은 과제 환경이 역동적인 변화한다는 것 자체이다. 따라서 이 정보가 행동 선택에 이용되기 위해서는 과제 중 경험하게 되는 역동적인 변화와 관련 개별 사건들을 매 시행마다 지속적으로 일화 기억으로 저장하여 여러 정보들과 통합되어야 할 가능성이 높다. 즉, 본 연구 결과는 스트레스가 확률적 반전 학습 과제 probabilistic reversal learning task 수행 시, 해마 hippocampus 의 일화기억 episodic memory 기능에 부정적인 영향을 미침으로써 행위자의 유연한 행동 대처를 방해함을 시사한다. 보상을 기반으로 하는 의사결정 과제 수행 중, 스트레스가 해마 기능을 방해하는 것으로 행위자의 적응적 의사결정 행동을 손상시킨다는 것이, 본 연구 결과에서의 중요한 시사점이라고 할 수 있겠다.

IV. 종합 논의

스트레스가 우리의 뇌에 구조적이나 기능적으로 미치는 영향에 대해서는 많은 연구가 진행되어 왔으나, 의사결정의 과정 및 결과에 대해 어떤 영향을 어떻게 미치는지는 아직 명확하지 않다. 이에 본 연구에서는 스트레스가 보상을 기반으로 하는 의사결정 행동에 어떤 영향을 미치는지 인지행동적 및 신경활동적 수준에서 다각적으로 탐색했다. 연구 결과, 스트레스는 강화학습의 여러 요소에 관여함으로써 행동 선택을 조절할 뿐 아니라, 스트레스의 정도에 따라 그 효과의 방향성이 다르다는 것이 확인되었다. 연구 결과를 바탕으로 다음과 같이 세 가지 시사점을 제안한다.

첫째, 스트레스는 행위자_{agent}가 행동을 선택할 때 목표지향적 또는 모형 기반_{model-based} 체계의 역할을 감소시키고, 습관적 또는 모형 부재_{model-free} 체계의 역할을 증가시키며, 스트레스로 인한 두 행동 체계의 변화는 서로 독립적일 가능성이 있다. 특히 연구 1에서 스트레스는 모형 부재_{model-free} 행동 양상 중, 보상을 받지 못한 행동을 다른 행동으로 전환하는 모형 부재_{model-free}의 손실-전환_{loss-switch} 양상을 증진시켰고, 이는 뇌에서 일어나는 행동의 긍정적 결과 처리와 부정적 결과 처리 과정에 스트레스가 다른 영향을 미칠 가능성을 시사한다.

둘째, 스트레스는 역동적으로 변화하는 환경에서 유연한 행동 선택을 방해한다. 즉, 스트레스는 해마_{hippocampus}의 역할, 즉 행위자

자신이 선택한 행동의 기대치를 계산하여 추적하는 기능을 방해하는 것으로 나타났으며, 이는 과제의 추상적인 규칙을 파악하는 것을 방해하여 결과적으로 적응적인 행동 선택에 부정적인 영향을 미치는 것으로 보인다. 또한 스트레스 처치 조건에서는 학습률_{learning rate}의 저하가 나타났으며, 이는 특히 목표지향적인 행동이 더욱 요구되는 상황에서 두드러졌다. 선택한 행동의 기대치_{expected value}와 실제 가치 간의 차이는 다음 시행의 행동 선택 시 참고 정보가 되며, 역동적인 상황에서는 매 시행마다 이 정보의 역할이 더욱 중요하다. 강화학습_{reinforcement learning} 모형에서 학습률_{learning rate}은 이 정보의 가중치로, 낮을 경우에는 이 정보가 이후 행동에 반영되는 정도가 저조하다. 본 연구 결과는 스트레스가 학습률_{learning rate}을 저하시킴으로써 변화하는 환경에서 유연한 행동 선택을 방해할 가능성을 시사한다.

셋째, 스트레스는 처치 수준에 따라 행동 선택에 양방향의 서로 다른 영향력을 행사하며, 이러한 현상에는 대뇌 내측 전전두엽_{medial prefrontal cortex}, 상측 측두엽_{superior temporal cortex}의 의사결정 관련 신경활성화 변화가 기저한다. 이 두 영역은 경미한 스트레스에는 신경활성화가 증가하며 목표지향적 행동이 증가하나, 스트레스가 가중되면 신경활성화 수준이 감소하고 이는 목표지향적 행동 감소와 동반된다. 즉, 내측 전전두엽_{medial prefrontal cortex}과 상측 측두엽_{superior temporal cortex}은 스트레스와 행동 선택 간 비단조적 관계에 대해 신경활동 수준의 조절자 역할을 하는 것으로 보이며, 특히 내측 전전두엽_{medial prefrontal cortex}은 스트레스의

수준에 따라 목표지향 체계뿐 아니라 습관적 행동 체계의 영향력을 조절하는 것으로 보인다. 본 연구 결과는 내측 전전두엽^{medial prefrontal cortex}, 특히 복측 영역^{ventromedial prefrontal cortex}이 스트레스의 양방향적 영향으로 인해 목표지향적 행동 체계와 습관 행동 체계 간 균형을 변화시키는 주요 영역일 가능성을 시사한다.

본 연구의 제한점은 다음과 같다. 연구 1에서 나타난 스트레스 관련 학습률^{learning rate} 저하나, 부적 피드백 후 습관행동의 증가 양상이 연구 2에서 재현되지 않았고, 그로 인해 두 스트레스 영향과 관련된 신경활성화 양상은 탐색할 수 없었다. 이는 두 실험 환경이 달라서 연구 2의 스트레스 무처치 조건의 피험자들의 과제 수행이 연구 1의 통제 조건 피험자들의 수행과 달랐기 때문으로 보인다. 연구 1에서는 모든 피험자들이 낮선 실험실에 혼자 앉아서 과제 수행을 한 것에 비해, 연구 2의 무처치 조건의 피험자들이 오후 1시 이후 따뜻한 물에 손을 담그는 통제 실험 처치를 받고 뇌 영상 스캐너에 누워 과제를 수행함으로써, 수행 중 이루어져 각성 수준이 저하된 것이 그 원인이라 생각된다. 또한 연구 2에서는 조건 별 피험자수가 적었고, 특히 의사결정 과제가 연구 1보다 시행 수가 적었던 것(fMRI 실험 시, 스트레스 반응의 지속 시간을 고려하여 연구 1에 비해 과제의 시행 수를 감소시킴). 이 실험 결과의 통계적 유의성 검증을 어렵게 한 것으로 보인다. 스트레스 반응에는 개인차가 있을 뿐 아니라, 스트레스가 주어진 상황이나 그 이후의 상황에 의해서도 영향이 달라질 수 있다. 이에 추후 연구들에서

본 연구의 결과가 확인 및 검증되는 것이 필요하며, 특히 역동적인 환경에서의 의사결정에 있어서 스트레스가 학습률(learning rate)를 저하시키는 현상에 대한 신경활성화 수준의 탐색을 후속 연구로 제안한다.

본 연구는 스트레스가 의사결정에 미치는 영향을 탐색하여 그 인지행동적 영향과 기저의 신경활동적 현상을 밝혔고, 스트레스 정도에 따른 양면적인 영향을 확인했다. 스트레스는 부적응적인 습관 행동을 촉진하며 중독 재발의 주요한 위험 요소로서 작용하나(Le et al., 2005; Koob and Kreek, 2007; Cleck and Blendy, 2008; Sinha, 2008), 아직 이러한 현상의 기전을 파악하는 연구가 부족하고, 관련 임상군을 진단 및 평가하는 방법이 제시된 바 없다. 이에 본 연구 결과는 의사결정에 있어서, 일상적인 문제에서부터 스트레스 관련 병리적 문제까지, 스트레스가 미치는 다양한 영향을 이해하는 데에 중요한 역할을 할 것으로 보인다. 또한 본 연구에서 개발된 2 단계 반전학습 과제는, 예상치 못하게 변화하는 환경에서의 의사결정자의 강화학습 행동 특성을 계량화하는 효과적인 패러다임으로 확인되었으므로, 추후 임상 및 연구 장면에서 부적응적인 의사결정 행동을 보이는 중독, 강박증, 섭식 장애 등 다양한 임상군의 진단 및 평가에 활용될 것으로 기대된다.

참고문헌

- 백영묘 (2010) 지각된 스트레스: 척도 표준화 및 기억 기능에 대한 영향 탐색. In: 심리학과: 서울대학교 대학원.
- Adam TC, Epel ES (2007) Stress, eating and the reward system. *Physiology & Behavior* 91:449-458.
- Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JD (2006) Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50:507-517.
- Anderson CR (1976) Coping behaviors as intervening mechanisms in the inverted-U stress-performance relationship. *Journal of Applied Psychology* 61:30.
- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407-419.
- Balleine BW, Dickinson A (2005) Effects of outcome devaluation on the performance of a heterogeneous instrumental chain. *International journal of comparative psychology* 18.
- Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nature neuroscience* 7:404-410.

- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nature neuroscience* 10:1214–1221.
- Bernacchia A, Seo H, Lee D, Wang XJ (2011) A reservoir of time constants for memory traces in cortical neurons. *Nature neuroscience* 14:366–372.
- Bromberg–Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* 68:815–834.
- Cahill L, Alkire MT (2003) Epinephrine enhancement of human memory consolidation: Interaction with arousal at encoding. *Neurobiology of learning and memory* 79:194–198.
- Cannon WB (1929) Vob. IX JULY, 1929 No. 3 ORGANIZATION FOR PHYSIOLOGICAL HOMEOSTASIS. *Physiological reviews* 9.
- Cleck JN, Blendy JA (2008) Making a bad thing worse: adverse effects of stress on drug addiction. *The Journal of clinical investigation* 118:454–461.
- Cohen S, Williamson GM (1988) Perceived stress in a probability sample of the United states In: *the social psychology of health: claremont symposium on applied social psychology* (Spacapan S, Oskamp S, eds). Newbury Park, CA: Sage.

- Daw ND (2011) Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII* 23:3–38.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–1215.
- De Kloet ER, Joëls M, Holsboer F (2005) Stress and the brain: from adaptation to disease. *Nature Reviews Neuroscience* 6:463–475.
- Diamond DM, Bennett MC, Fleshner M, Rose GM (1992) Inverted-U relationship between the level of peripheral corticosterone and the magnitude of hippocampal primed burst potentiation. *Hippocampus* 2:421–430.
- Dias-Ferreira E, Sousa JC, Melo I, Morgado P, Mesquita AR, Cerqueira JJ, Costa RM, Sousa N (2009) Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* 325:621–625.
- Dickerson KC, Li J, Delgado MR (2011) Parallel contributions of

- distinct human memory systems during probabilistic learning. *Neuroimage* 55:266–276.
- Dickinson A (1985) Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 308:67–78.
- Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology* 22:1075–1081.
- Domes G, Rothfischer J, Reichwald U, Hautzinger M (2005) Inverted-U function between salivary cortisol and retrieval of verbal memory after hydrocortisone treatment. *Behavioral neuroscience* 119:512.
- Ferragud A, Haro A, Sylvain A, Velazquez-Sanchez C, Hernandez-Rabaza V, Canales J (2010) Enhanced habit-based learning and decreased neurogenesis in the adult hippocampus in a murine model of chronic social stress. *Behavioural brain research* 210:134–139.
- Flier JS, Underhill LH, McEwen BS (1998) Protective and damaging effects of stress mediators. *New England Journal of Medicine* 338:171–179.
- Foerde K, Shohamy D (2011) Feedback timing modulates brain

systems for learning in humans. *The Journal of Neuroscience* 31:13157–13167.

Gläscher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action–based value signals during reward–related decision making. *Cerebral cortex* 19:483–495.

Gold P, Van Buskirk R, McGaugh J (1975) Effects of hormones on time–dependent memory storage processes. *Progress in brain research* 42:210.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience* 28:5623–5630.

Henckens MJ, Hermans EJ, Pu Z, Joëls M, Fernández G (2009) Stressed memories: how acute stress affects memory formation in humans. *The Journal of Neuroscience* 29:10111–10119.

Higgins RL, Marlatt GA (1975) Fear of interpersonal evaluation as a determinant of alcohol consumption in male social drinkers. *Journal of abnormal psychology* 84:644.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error

in the temporal prediction of reward during learning. *Nature neuroscience* 1:304–309.

Howard RA (1960) DYNAMIC PROGRAMMING AND MARKOV PROCESSES. In. USA: Massachusetts Institute of Technology Press.

Hull C (1943) *Principles of behavior*.

Hull JG, Young RD (1983) Self-consciousness, self-esteem, and success-failure as determinants of alcohol consumption in male social drinkers. *Journal of Personality and Social Psychology* 44:1097.

Ito M, Doya K (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *The Journal of neuroscience* 29:9861–9874.

Jensen J, Smith AJ, Willeit M, Crawley AP, Mikulis DJ, Vitcu I, Kapur S (2007) Separate brain regions code for salience vs. valence during reward prediction in humans. *Human brain mapping* 28:294–302.

Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS biology* 4:e233.

Kim J, Diamond DM (2002) The stressed hippocampus, synaptic

plasticity and lost memories. *Nature Reviews Neuroscience* 3:453–462.

Kirschbaum C, Kudielka BM, Gaab J, Schommer NC, Hellhammer DH (1999) Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus–pituitary–adrenal axis. *Psychosomatic medicine* 61:154–162.

Koob G, Kreek M (2007) Stress, dysregulation of drug reward pathways, and the transition to drug dependence. *American Journal of Psychiatry* 164:1149–1159.

Le A, Harding S, Juzysch W, Funk D, Shaham Y (2005) Role of alpha-2 adrenoceptors in stress-induced reinstatement of alcohol seeking and alcohol self-administration in rats. *Psychopharmacology* 179:366–373.

Lee D (2013) Decision making: from neuroscience to psychiatry. *Neuron* 78:233–248.

Lee H, Ghim J–W, Kim H, Lee D, Jung M (2012) Hippocampal neural correlates for values of experienced events. *The Journal of Neuroscience* 32:15053–15065.

Lee Sang W, Shimojo S, O’ Doherty John P (2014) Neural Computations Underlying Arbitration between Model–Based and Model–free Learning. *Neuron* 81:687–699.

- Liston C, Miller MM, Goldwater DS, Radley JJ, Rocher AB, Hof PR, Morrison JH, McEwen BS (2006) Stress-induced alterations in prefrontal cortical dendritic morphology predict selective impairments in perceptual attentional set-shifting. *The Journal of Neuroscience* 26:7870–7874.
- Lupien SJ, Lepage M (2001) Stress, memory, and the hippocampus: can't live with it, can't live without it. *Behavioural brain research* 127:137–158.
- McEwen BS (2000) Allostasis and allostatic load: implications for neuropsychopharmacology. *Neuropsychopharmacology* 22:108–124.
- McEwen BS, Sapolsky RM (1995) Stress and cognitive function. *Current opinion in neurobiology* 5:205–216.
- Mendelson JH, Sholar MB, Goletiani N, Siegel AJ, Mello NK (2005) Effects of low- and high-nicotine cigarette smoking on mood states and the HPA axis in men. *Neuropsychopharmacology* 30:1751–1763.
- Mitra R, Sapolsky RM (2008) Acute corticosterone treatment is sufficient to induce anxiety and amygdaloid dendritic hypertrophy. *Proceedings of the National Academy of Sciences* 105:5573–5578.

- Mizoguchi K, Yuzurihara M, Ishige A, Sasaki H, Chui D-H, Tabira T (2000) Chronic stress induces impairment of spatial working memory because of prefrontal dopaminergic dysfunction. *The Journal of Neuroscience* 20:1568–1574.
- Murphy B, Arnsten A, Goldman-Rakic P, Roth R (1996) Increased dopamine turnover in the prefrontal cortex impairs spatial working memory performance in rats and monkeys. *Proceedings of the National Academy of Sciences* 93:1325–1329.
- O'Connor T, O'halloran D, Shanahan F (2000) The stress response and the hypothalamic-pituitary-adrenal axis: from molecule to melancholia. *Qjm* 93:323–333.
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
- O'Doherty JP, Hampton A, Kim H (2007) Model-Based fMRI and Its Application to Reward Learning and Decision Making. *Annals of the New York Academy of Sciences* 1104:35–53.
- Otto AR, Gershman SJ, Markman AB, Daw ND (2013a) The Curse of Planning Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological*

science 24:751–761.

Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND (2013b) Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America* 110:20941–20946.

Paulus MP, Feinstein JS, Leland D, Simmons AN (2005) Superior temporal gyrus and insula provide response and outcome-dependent information during assessment and action selection in a decision-making situation. *Neuroimage* 25:607–615.

Penner MR, Mizumori SJ (2012) Neural systems analysis of decision making during goal-directed navigation. *Progress in neurobiology* 96:96–135.

Pomerleau CS, Pomerleau OF (1987) The effects of a psychological stressor on cigarette smoking and subsequent behavioral and physiological responses. *Psychophysiology* 24:278–285.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory* 2:64–99.

Roosendaal B, McEwen BS, Chattarji S (2009) Stress, memory and the amygdala. *Nature Reviews Neuroscience* 10:423–433.

- Salehi B, Cordero MI, Sandi C (2010) Learning under stress: the inverted-U-shape function revisited. *Learning & Memory* 17:522–530.
- Sapolsky RM (1996) Why stress is bad for your brain. *Science* 273:749–750.
- Sauro MD, Jorgensen RS, Teal Pedlow C (2003) Stress, glucocorticoids, and memory: a meta-analytic review. *Stress: The International Journal on the Biology of Stress* 6:235–245.
- Schönberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *The Journal of Neuroscience* 27:12860–12867.
- Schneiderman N, Ironson G, Siegel SD (2005) Stress and health: psychological, behavioral, and biological determinants. *Annual Review of Clinical Psychology* 1:607.
- Schultz W (1998) Predictive reward signal of dopamine neurons. *Journal of neurophysiology* 80:1–27.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Schwabe L, Wolf OT (2009) Stress prompts habit behavior in humans. *The Journal of Neuroscience* 29:7191–7198.

- Schwabe L, Dalm S, Schächinger H, Oitzl MS (2008) Chronic stress modulates the use of spatial and stimulus–response learning strategies in mice and man. *Neurobiology of learning and memory* 90:495–503.
- Schwabe L, Tegenthoff M, Höffken O, Wolf OT (2012) Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal–directed action in the human brain. *The Journal of Neuroscience* 32:10146–10155.
- Schwarz G (1978) Estimating the dimension of a model. *The annals of statistics* 6:461–464.
- Selye H (1946) THE GENERAL ADAPTATION SYNDROME AND THE DISEASES OF ADAPTATION 1. *The Journal of Clinical Endocrinology & Metabolism* 6:117–230.
- Selye H (1976) *Stress in health and disease*: Butterworths Boston.
- Simon DA, Daw ND (2011) Environmental statistics and the trade–off between model–based and TD learning in humans. *Advances in Neural Information Processing Systems* 24:127–135.
- Sinha R (2008) Chronic stress, drug use, and vulnerability to addiction. *Annals of the New York Academy of Sciences* 1141:105–130.

- Spence KW (1956) Behavior theory and conditioning.
- Sutton RS, Barto AG (1998) Introduction to reinforcement learning:
MIT Press.
- Thorndike EL (1911) Animal intelligence: Experimental studies:
Macmillan.
- Tolman EC (1951) Purposive behavior in animals and men: Univ of
California Press.
- Ulrich-Lai YM, Herman JP (2009) Neural regulation of endocrine
and autonomic stress responses. Nature Reviews
Neuroscience 10:397–409.
- Wimmer GE, Daw ND, Shohamy D (2012) Generalization of value in
reinforcement learning by humans. European Journal of
Neuroscience 35:1092–1104.
- Yerkes RM, Dodson JD (1908) The relation of strength of stimulus
to rapidity of habit-formation. Journal of comparative
neurology and psychology 18:459–482.

Abstract

The effects of stress on decision–making

Heyeon Park

Department of Psychology

The Graduate School

Seoul National University

Previous computational and lesion studies suggested that behaviors might be controlled by at least two partially separate neuroanatomical systems. Habit or model–free reinforcement learning incrementally shapes behaviors by trial and error, whereas goal–directed behaviors or model–based reinforcement learning rely on the internal model of the decision maker’s environment and motivation. Stress appears to influence the arbitration between a goal–directed and a habit processing during decision making. However, it is unclear how stress might modulate the arbitration between them. In addition, although too much stress impairs performance in a variety of tasks,

it can also improve cognitive performance when a certain level of arousal is beneficial, as often referred to as the Yerkes–Dodson law (inverted U). However, whether and how such bidirectional effects of stress manifest as a potentially shifting balance between habit versus goal–directed systems has not been investigated.

Here, I examined the multiple effects of stress on model–free and model–based decision making at both behavioral and neural levels. Firstly, in Study 1, I tested whether stress biases the contribution of model–free and model–based reinforcement learning processes and alters the rate of incorporating new information from the environment. Participants were randomly assigned to stress and control conditions, and performed a two–stage reversal learning task in which the reward probabilities underwent reversals without notice. By applying computational models to participants’ choices, I quantified the extent to which choices were influenced by model–free and model–based processing and the degree to which choices were updated with new information, the learning rate. The results showed that under stress, the contribution of model–based RL to decision–making was diminished while the tendency to switch after negative feedback as predicted by model–free RL was enhanced. Moreover, the size of these two effects varied independently across

different participants. Also, the learning rate, at which decision variables in RL models are updated by new experience, was decreased due to stress.

Secondly, in the study 2, I examined the dose-dependent effects of stress on model-free and model-based reinforcement learning, using computational models and functional magnetic resonance imaging. Participants were randomly assigned to one of the three conditions, no-stress, single-stress-treatment, and double-stress-treatment. Two types of psychological stress protocol were employed in the study, which was immediately followed by a two-stage Markov decision-making task in which the reward probabilities for the choices in the second stage underwent reversals without notice. Analogous to the Yeskes-Dodson hypothesis, the parameter estimates representing goal-directed behaviors increased in a single-stress-treatment condition but decreased in a double-stress-treatment condition. Similarly, BOLD signals in the medial prefrontal and superior temporal gyrus during decision making were accord with the inverted-U shaped behavioral results. Also, the chosen value related signals in the right hippocampus were decreased in the two stressed conditions, which was related to the impairment of reversal learning.

In this study, I suggest that first, stress can play a critical role in arbitrating the two controllers (habit and goal-directed) of decision making and second, the effects of stress on decision making might be bidirectional depending on the level of stress currently experienced by the decision-maker, and finally, the role of hippocampus in adaptive behavior in a changing environment be impaired under stress. These results have implications for the underlying neural mechanism mediating the effects of stress on the formation of maladaptive habits, such as addictive behavior, as well as dysfunctional behaviors associated with stress-related disorders.

Keywords : acute stress, reinforcement learning, reversal learning, computational models, fMRI

Student Number: 2007-30052