

# Fonksiyonel Bilinç Çöküş Teorisi (FBÇT)

Functional Consciousness Collapse Theory (FCCT)

Muhammed Yasin Özkaya

Kasım 2025

## Abstract

Fonksiyonel Bilinç Çöküşü Kuramı (FBÇT), bilinci metafizik bir varlık olarak değil, durumsal bilgiden türeyen hesaplanabilir bir seçim-dinamiği olarak ele alan bir çerçeve sunar. Kuram, her bilinç anının üç temel bileşenin etkileşimiyle belirlendiğini varsayar: duyuşal durum  $S_t \in \mathbb{R}^{n_S}$ , hafıza yapısı  $M_t \in \mathbb{R}^{n_M}$  ve öznenin değer ağırlıklarını temsil eden önem vektörü  $W_t \in \Delta^{k-1}$ . Bu üçlü tarafından belirlenen bilişsel durum uzayı  $X$ , kuramın temel dinamiklerinin gerçekleştiği içsel temsil alanıdır.

FBÇT’de bilinç durumu  $C_t$ ,  $(S_t, M_t, W_t)$  üçlüsünün oluşturduğu bir olasılık çekirdeği  $\mu_t(\cdot)$  üzerinden üretilen aday durumların bir skor fonksiyonu  $f$  ile değerlendirilmesi ve olasılıksal bir seçim mekanizmasıyla çökmesiyle elde edilir. Bu süreç matematiksel olarak  $\pi_t(x) \propto \exp(\beta_t f(x, S_t, M_t, W_t))$  şeklindeki bir politika dağılımıyla ifade edilir ve  $C_t \sim \pi_t$  olacak biçimde gerçekleşen bir seçim çöküşüyle sonuçlanır. Bilinç düzeyi, ortaya çıkan politikanın arka plan dağılımdan ne ölçüde ayrıştığını ölçen  $L_t = D_{KL}(\pi_t \| P_t)$  metriğiyle nicelenir; bu nicelik, dikkatin toplanması, odaklanma ve karar belirginliği ile doğrudan ilişkilidir.

Kuram, öznel deneyim (qualia) problemini,  $X \times S$  üzerinde tanımlanan bir eşdeğerlik ilişkisi ile formelleştirir. Buna göre iki durum  $(x_1, S_1)$  ve  $(x_2, S_2)$  öznenin tüm fonksiyonel çıktıları açısından ayırt edilemiyorsa aynı qualia sınıfına aittir. Bu yapı, qualia uzayını  $Q = (X \times S) / \sim$  şeklindeki bir bölüm uzayı olarak tanımlar. Deneysel olarak kullanılan tersine mühendislik yaklaşımı olan  $\Phi : X \times S \rightarrow \mathbb{R}^m$ , bu soyut uzayın bir koordinat sistemini oluşturur;  $\Phi$  öğrenilebilir bir parametrisasyondur ve qualia eşdeğerlik sınıflarının ampirik olarak belirlenmesini sağlar.

FBÇT, böylece homunculus problemini ortadan kaldırarak, öznel niteliği fonksiyonel eşdeğerlik üzerinden tanımlayan ve özgür iradeyi hafıza ve değer dinamiklerinin güncellenebilirliği çerçevesinde yeniden yorumlayan bütünlüklü bir matematiksel model sunar. Kuram, aynı  $(S, M, W)$  yapılarına sahip iki sistemin aynı bilinç durumunu üretmesi gerektiğini öne sürer; bu öngörü nörobilimsel ölçümler, klinik gözlemler ve yapay bilinç sistemleri üzerinde sınanabilir niteliktedir.

Sonuç olarak FBÇT, bilinci gizemli bir fenomen olmaktan çıkararak, durum uzayları, seçim mekanizmaları ve eşdeğerlik yapıları üzerinden bilimsel olarak incelenebilir, hesaplanabilir ve biçimsel olarak modellenenebilir bir olgu hâline getirir.

## 1 Giriş

### 1.1 Bilinç Probleminin Bilimsel Konumu

Bilinç, modern bilimin hem en eski hem de en dirençli kavramsal sorunlarından biridir. İnsan beyninin biyofiziksel düzeyde incelenmesi, nöral popülasyonların kolektif dinamiklerini, büyük ölçekli bağlantısalılığı ve bilişsel işlevlerin hesaplamalı temellerini giderek daha ayırt edilebilir hâle getirmiştir. Bununla birlikte nesnel fiziksel süreçlerin nasıl olup da öznel bir deneyim alanı - bir bakış açısı, bir “ben” - ürettiği hâlen açık bir problem olarak durmaktadır [1–3]. Dolayısıyla bilinç araştırmaları, biyolojik mekanizmalar ile fenomenolojik içerik arasındaki ilişkinin nasıl tanımlanması gerektiği üzerine temellendirilen uzun süreli bir metodolojik gerilim taşımaktadır.

## 1.2 Paradigmatik Yaklaşımlar

Mevcut teoriler iki ana paradigma altında toplanabilir.

**Fonksiyonalist yaklaşımlar**, bilinci beynin bilgi işleme mimarisinin bir sonucu olarak değerlendirir. Global Workspace Theory [4, 5], Higher-Order Thought Theory [6] ve Attention Schema Theory [7] gibi modeller, öznel farkındalığı rekabet eden temsiller arasında küresel erişim, yüksek-dereceli izleme veya dikkat modellemesi gibi işlevsel mekanizmalar üzerinden açıklar. Bu yaklaşımlar deneysel olarak test edilebilir olmaları bakımından güçlüdür; ancak eleştirmenler, bu modellerin fenomenolojik içeriği - deneyimin “nasıl hissettirdiği” yönünü - yetersiz biçimde ele aldığını savunur [8].

Buna karşılık **non-fonksiyonalist yaklaşımlar** bilinci daha temel bir ontolojik kategori olarak kabul eder. Integrated Information Theory (IIT) [9, 10], bilinci  $\Phi$  ile ölçülen bütünleşmiş bilgi miktarı üzerinden tanımlayarak, hesaplamalı sistemlerden ziyade nedensel yapıya öncelik tanır. Kuantum tabanlı teoriler [11, 12] ise bilincin beyindeki kuantum süreçlerinden kaynaklandığını öne sürer. Bu teoriler fenomenolojiyi merkeze alırken, genellikle biyolojik verilerle sınırlı ilişki kurmaları veya deneysel olarak doğrulanabilir olmamaları nedeniyle eleştirilir.

## 1.3 Kolay ve Zor Problem Çerçevesi

Bu iki yaklaşım arasındaki temel ayrım, Chalmers tarafından tanımlanan *kolay problem* ve *zor problem* ayrımına dayanır [1]. Kolay problem, algı, dikkat, bellek ve karar verme gibi bilişsel işlevlerin mekanizmalarını açıklamayı içerir. Zor problem ise fiziksel süreçlerin neden ve nasıl öznel deneyim ürettiğine ilişkin sorudur. Fonksiyonalist modeller kolay probleme güçlü yanıtlar verirken zor probleme sınırlı yaklaşır; non-fonksiyonalist modeller ise fenomenolojiyi açıklamaya çalışırken hesaplanabilir mekanizmaları geri planda bırakır.

Bu makale, her iki yaklaşımın da paylaştığı merkezi bir açığı vurgular: Mevcut teorilerin çoğu ya bilinci salt fonksiyonel bir süreç olarak ele almakta ve fenomenolojiyi türevsel görmektedir, ya da fenomenolojiyi temel almakta ve bilişsel mekanizmayı belirsizleştirmektedir. Oysa bilinç, hem hesaplanabilir bir süreç hem de öznel bir deneyim olarak ortaya çıkar. Bu iki düzeyi bağlayan ortak bir matematiksel çerçeve, teorilerin çoğunun eksik bıraktığı yapısal bütünlüğü sağlayabilir.

## 1.4 FBÇT’nin Önerdiği Çerçeve

Bu bağlamda Fonksiyonel Bilinç Çöküşü Teorisi (FBÇT), bilinci üç temel bileşenin etkileşiminden doğan *olasılıksal bir seçim çöküşü* olarak yeniden tanımlar. Duyusal durum  $S_t$ , hafıza yapısı  $M_t$  ve önem vektörü  $W_t$  tarafından belirlenen bir olasılık çekirdeği  $\mu_t$ , içsel temsil uzayında aday bilişsel durumlar üretir; bu adaylar çok bileşenli bir değer fonksiyonu aracılığıyla değerlendirilir ve  $\pi_t(x) \propto \exp(\beta_t f(x, S_t, M_t, W_t))$  şeklinde tanımlanan bir politika dağılımı üzerinden çöker. Bu çöküş, bilinç anını oluşturan  $C_t$  durumunun seçilmesini sağlar.

FBÇT, fenomenolojik içeriği  $(x, S)$  çiftleri üzerinde tanımlanan bir eşdeğerlik ilişkisiyle matematiksel olarak yapılandırır. Fonksiyonel olarak ayırt edilemeyen durumların oluşturduğu  $Q = (X \times S)/\sim$  uzayı, qualia’nın soyut yapısını belirler. Bu soyut uzayın koordinatlandırılması, tersine mühendislik yoluyla öğrenilebilen  $\hat{\Phi} : X \times S \rightarrow \mathbb{R}^m$  haritalamasıyla sağlanır; böylece deneysel ölçümlerin fenomenolojik yapıyla ilişkilendirilmesi formelleşmiş olur.

FBÇT, homunculus problemini çöküş mekanizmasını emergent bir süreç olarak tanımlayarak, qualia problemini eşdeğerlik sınıfları üzerinden yeniden ifade ederek ve özgür iradeyi  $M_t$  ve  $W_t$ ’nin dinamik güncellenebilirliği bağlamında ele alarak felsefenin klasik sorunlarına fonksiyonel çözümler sunar. Ayrıca teoremin, aynı  $(S, M, W)$  yapılarına sahip iki sistemin aynı bilinç durumunu üretmesi gerektiği yönündeki tahmini, nörogörüntüleme verileri, klinik vakalar ve yapay bilinç modelleri üzerinden doğrudan test edilebilir.

## 1.5 Makalenin Yapısı

Bu makalenin geri kalanındaki yapı şu şekildedir: Bölüm 2 bilincin tarihsel arka planını ve mevcut teorilerin sistematik analizini sunar. Bölüm 3, FBÇT'nin matematiksel formalizmini tanımlar. Bölüm 5, klasik felsefi problemlere getirilen fonksiyonel çözümleri tartışır. Bölüm 6, teoremin test edilebilir öngörülerini ele alır. Bölüm 8, FBÇT'nin diğer teorilerle karşılaştırmasını ve sınırlamalarını tartışır. Son olarak Bölüm 9, kuramın geniş bilimsel bağlamdaki etkilerini özetler.

## 2 Arka Plan ve Literatür İncelemesi

Bilinç problemi, modern bilim felsefesinin en eski ve en tartışmalı konularından biridir. Bu bölümde, problemin tarihsel gelişimini, mevcut teorilerin temel yaklaşımlarını ve her birinin güçlü/zayıf yanlarını sistematik olarak inceleyeceğiz.

### 2.1 Bilinç Probleminin Tanımı ve Tarihçesi

#### 2.1.1 Kartezyen İkilik ve Zihin-Beden Problemi

Modern bilinç tartışmasının kökleri Descartes'ın (1641) *res cogitans* (düşünen şey) ve *res extensa* (uzamsal şey) ayırımına dayanır. Descartes, zihnin maddi olmayan bir töz olduğunu ve bedenle yalnızca pineal gland aracılığıyla etkileştiğini savunmuştur. Bu dualist yaklaşım, “zihin fiziksel dünyanın bir parçası mıdır?” sorusunu merkeze almış ve sonraki yüzyıllarda çeşitli yanıtlar üretilmiştir: materyalizm (zihin beynin bir ürünüdür), idealizm (sadece zihin gerçektir), ve paralelizm (ikisi birbirinden bağımsızdır).

20. yüzyılda, davranışçılık (behaviorism) bilinci tamamen göz ardı etmiş, sadece gözlemlenebilir davranışlara odaklanmıştır. Ancak 1950'lerden itibaren bilişsel devrim, zihinsel süreçlerin bilimsel araştırmaya konu olabileceğini göstermiştir. Bugün, bilinç nörobilim, bilişsel bilim, yapay zeka ve felsefenin kesiştiği interdisipliner bir alandır.

#### 2.1.2 Kolay Problem ve Zor Problem Ayrımı

Chalmers [1], bilinç araştırmasını iki kategoriye ayırmıştır:

##### Kolay Problemler.

Bilişsel işlevlerin mekanizmalarının açıklanması yani algı, dikkat, bellek, dil, karar verme gibi süreçler. “Kolay” demek, bu problemlerin basit olduğu anlamına gelmez; prensipte nörobilim ve bilişsel bilim yöntemleriyle çözülebilir oldukları anlamındadır.

##### Zor Problem.

Neden öznel deneyim var? Fiziksel süreçler nasıl oluyor da “bir şey hissettiriyor”? Örneğin, kırmızı dalga boyunu algılayan nöronlar var ama *kırmızının kırmızı hissi* neden oluşuyor? Bu fenomenolojik boyut, fonksiyonel açıklamalarla kapatılamaz gibi görünüyor.

Chalmers'a göre, hard problem *açıklayıcı uçurum* (explanatory gap) yaratır: Beynin fiziksel durumlarını tam bilsek bile, bunların neden belirli öznel deneyimler ürettiğini açıklayamayız. Bu, eliminativist (fenomenoloji bir yanılsamadır) ya da dualist (bilinç fiziksel değildir) yaklaşımlara yol açar.

Ancak bazı filozoflar bu ayrımı reddeder. Dennett [3], hard problem'in “gerçek bir problem olmadığını”, sadece sezgilerimizden kaynaklanan bir illüzyon olduğunu savunur. Ona göre, bilinç tamamen fonksiyonel açıklamalarla anlaşılabilir; qualia gibi kavramlar bilimsel açıklamada gereksizdir.

## 2.2 Mevcut Bilinç Teorileri

### 2.2.1 Global Workspace Theory (GWT)

#### Temel iddia.

Baars [4] ve Dehaene & Naccache [5] tarafından geliştirilen Global Workspace Theory (GWT), bilinci beynin bir “global yayın mekanizması” olarak tanımlar. Modele göre, beyin paralel çalışan çok sayıda özelleşmiş modülden oluşur (görsel işleme, işitsel işleme, motor kontrol, vb.). Bu modüller bilinçsiz çalışır; ancak belirli bir bilgi parçası “global workspace”e (küresel çalışma alanına) eriştiğinde tüm sistemde yayılır ve bilinçli hale gelir. Matematiksel olarak:

$$C_t = \begin{cases} 1 & \text{eğer } I_t \in \text{Global Workspace} \\ 0 & \text{aksi halde} \end{cases} \quad (1)$$

Burada  $I_t$ ,  $t$  anındaki bilgi parçasını;  $C_t$  ise bilinçli olup olmadığını gösterir.

#### Nörobiyolojik temel.

GWT, prefrontal ve parietal korteksin geniş ağ bağlantılarına dayanır. Bilinçli algı sırasında frontoparietal ağda yüksek frekans (gamma) senkronizasyonu gözlemlenir [13]; bilinçsiz işlemler ise yerel, kısa süreli aktivasyonlarla sınırlıdır.

#### Güçlü yanlar. • Nörogörüntüleme bulgularıyla uyumlu

- Test edilebilir tahminler üretiyor (örn. dikkat, çalışma belleği kapasitesi)
- Bilinçli vs. bilinçsiz işlemlemeyi ayırt edebiliyor
- Yapay sistemlere uygulanabilir

#### Zayıf yanlar. • Qualia problemini açıklamıyor: “global yayın” neden öznel deneyim üretiyor?

- “Workspace” metaforunun sınırları belirsiz
- Bilinç için gerek ve yeter koşul mu, yoksa eşlik eden durum mu sorusunu açık bırakıyor
- Fenomenolojik zenginliği (örn. duygu, estetik deneyim) sınırlı şekilde ele alıyor

### 2.2.2 Integrated Information Theory (IIT)

#### Temel iddia.

Tononi [9, 10] tarafından geliştirilen Integrated Information Theory (IIT), bilinci matematiksel bir ölçütle tanımlar:  $\Phi$  (phi).  $\Phi$ , sistem parçalarının birbirinden ne kadar fazla bilgi taşıdığı ve bu bilginin ne kadar indirgenemez olduğunu ölçer:

$$\Phi = \min_{S=S_1 \cup S_2} \text{KL}(p(S) \| p(S_1) \times p(S_2)). \quad (2)$$

$\Phi > 0$  ise sistem bilinçlidir;  $\Phi$  ne kadar büyükse bilinç o kadar “zengin”dir.

#### Ana çıkarımlar. • Her $\Phi > 0$ olan sistem bilinçlidir (panpsychism’e yaklaşır).

- Beyinde talamo-kortikal sistem yüksek  $\Phi$  değerine sahiptir.
- Serebellum düşük  $\Phi$  değerine sahiptir (çok nöron olmasına rağmen entegrasyon zayıftır).

#### Güçlü yanlar. • Matematiksel kesinlik ve fenomenolojiyi önceleme

- Bilinç “derecelerini” nicel olarak modelleme
- Klinik durumlarda (örn. vegetatif durum) test edilebilir öngörüler

#### Zayıf yanlar. • $\Phi$ hesaplaması NP-hard; büyük sistemler için pratik değil

- Panpsychism’i çağrıştıran çıkarımlar tartışmalı

- Nörobiyolojik mekanizmalarla bağ zayıf ve soyut
- “Entegre bilgi”nin neden öznel deneyim ürettiği açıklanmıyor
- Ampirik testler sınırlı ve dolaylı

### 2.2.3 Higher-Order Thought (HOT) Theory

#### Temel iddia.

Rosenthal [6] ve diğerlerine göre bilinç, “yüksek-düzey düşünceler” (higher-order thoughts) gerektirir. Bir zihinsel durumun bilinçli olması için o durum hakkında meta-kognitif bir temsil bulunmalıdır:

$$\text{Bilinli}(M) \iff \exists \text{HOT}(M). \quad (3)$$

Örneğin acı hissi bilinçlidir çünkü “acı hissettiğimin farkındayım” düşüncesi eşlik eder.

#### Güçlü yanlar. • Öz-farkındalık ve meta-bilişi açıklama kapasitesi

- Bilinçli/bilinçsiz ayrımını kavramsal olarak netleştirmesi
- Prefrontal hasar gibi nöropsikolojik bulgularla uyum

#### Zayıf yanlar. • Sonsuz regresyon riski: HOT’un kendisi bilinçli olmalı mı? (HOT-HOT-HOT...?)

- Bazı bilinç deneyimleri meta-kognitif olmayabilir (örn. ani bir ağrı)
- Qualia’yı açıklamıyor; HOT neden “bir şey hissettiriyor?”

### 2.2.4 Attention Schema Theory (AST)

#### Temel iddia.

Graziano [7, 14] bilinçli deneyimin beyin dikkat süreçlerini modellemesinden kaynaklandığını öne sürer. Beyindeki “attention schema”, dikkat kaynaklarının basitleştirilmiş bir modelidir ve bu modelin yan etkisi olarak “öznel deneyim” hissederiz; bilinç algısı esasında bir yanılsamadır.

#### Analoji.

Beyin, kol hareketleri için bir “body schema” tutar; benzer şekilde dikkat için bir “attention schema” üretir. Bilinç, bu şemanın yanlış atfıdır (“burada bir şey var”).

#### Güçlü yanlar. • Yapay sistemlere kolayca uygulanabilir

- Nörobiyolojik olarak makul (temporoparietal korteksle ilişkilendirilir)
- Illusion-based yaklaşım sayesinde qualia’yı “ayrı bir varlık” olarak varsaymaz

#### Zayıf yanlar. • Eliminativist yönelimi öznel deneyimi “gerçek değil” diye reddeder

- Sezgisel olarak ikna edici değildir; fenomenolojik veriyi değersizleştirir
- Eğer bilinç yanılsama ise, bu yanılsamayı deneyimleyen kim sorusunu yanıtsız bırakır

### 2.2.5 Predictive Processing / Free Energy Principle

#### Temel iddia.

Friston [15] ve Clark [16], beyin bir “Bayesian tahmin makinesi” olduğunu savunur. Beyin sürekli duyuşal girdileri tahmin eder ve tahmin hatasını (free energy) minimize etmeye çalışır:

$$F = -\log p(s|m), \quad (4)$$

burada  $s$  duyuşal girdi,  $m$  ise iç modeldir. Amaç  $F$ ’yi minimize etmektir.

### Bilinç yorumu.

Yüksek-düzey tahminlerin entegrasyonu bilinçli deneyimi oluşturur. Seth [17] için bilinç “kontrollü halüsinasyon”dur; dış dünyayı değil, beynin iç modelini deneyimleriz.

**Güçlü yanlar.** • Geniş nörobiyolojik destek ve bütüncül çerçeve (algı-eylem-öğrenme)

- Interosepsiyon ile bilinç arasındaki bağı açıklayabilmesi
- Psikedelik deneyimler, dikkat dinamikleri gibi fenomenleri tek modelde açıklama potansiyeli

**Zayıf yanlar.** • Aşırı genel; yanlışlanabilirliği tartışmalı

- Bilinç için özel bir mekanizma tanımlamıyor, “yüksek seviye tahmin” ifadesinde kalıyor
- Qualia problemi açık: tahmin hatası neden belirli bir his yaratıyor?

### 2.2.6 Quantum Consciousness Theories

#### Temel iddia.

Penrose [11] ve Hameroff [12], bilincin kuantum süreçlerinden kaynaklandığını öne sürer. Özellikle nöronların mikrotübüllerinde “orchestrated objective reduction” (Orch-OR) gerçekleşir ve kuantum dalga fonksiyonu çöküşü bilinçli deneyimi üretir.

**Motivasyon.** • Klasik hesaplama Gödel’in eksiklik teoremi nedeniyle bilinci açıklayamaz

- Öznel deneyim kuantum ölçüm problemine benzer
- Mikrotübüller yeterince izole ve soğuk olabilir (iddia edilen şekilde)

**Güçlü yanlar.** • Bilinç için “özel” bir fiziksel süreç önermesi

- Qualia problemine radikal, fizik temelli bir çözüm sunma iddiası

**Zayıf yanlar.** • Nörobiyolojik destek son derece zayıf; mikrotübüllerde kuantum coherence kanıtlanmadı

- Gödel argümanı tartışmalı ve ikna edici bulunmuyor
- Test edilemez veya mevcut teknolojide doğrulanamaz
- Bilim camiasında geniş kabul görmüyor

### 2.3 Teorilerin Karşılaştırmalı Analizi

Tablo 1’de mevcut bilinç teorilerinin temel özellikleri karşılaştırılmıştır.

Table 1: Mevcut bilinç teorilerinin karşılaştırması

Teori	Matematiksel	Test edilebilir	Qualia	Nörobiyoloji
GWT	Orta	Yüksek	Düşük	Yüksek
IIT	Yüksek	Düşük	Orta	Orta
HOT	Düşük	Orta	Düşük	Orta
AST	Orta	Yüksek	Düşük (eliminativist)	Yüksek
Predictive Processing	Yüksek	Orta	Orta	Yüksek
Kuantum	Düşük	Çok Düşük	Yüksek (?)	Çok Düşük
FBÇT (bu çalışma)	Yüksek	Yüksek	Orta-Yüksek	Yüksek

## 2.4 Literatürdeki Boşluk

Yukarıdaki inceleme, mevcut teorilerin ortak bir eksiğini ortaya koymaktadır: Teoriler ya *fonksiyonel* (GWT, HOT, AST, PP) ya da *fenomenolojik* (IIT, Kuantum) kutuplarda konumlanıyor. Fonksiyonel teoriler test edilebilir ama qualia'yı açıklayamıyor; fenomenolojik teoriler qualia'ya öncelik veriyor ama mekanizma belirsiz ve test edilemez.

İhtiyaç duyulan şey, **hem fonksiyonel hem de fenomenolojik boyutu birleştiren, matematiksel olarak kesin, test edilebilir ve nörobiyolojik olarak makul bir çerçevedir.** Fonksiyonel Bilinç Çöküşü Teorisi (FBÇT), bu boşluğu doldurmayı amaçlar.

FBÇT'nin temel yeniliği şudur: Bilinci statik bir durum değil, *dinamik bir süreç* - özellikle bir "seçim çöküşü" - olarak tanımlar. Bu yaklaşım, qualia'yı fonksiyonel bileşenlerin ( $S$ ,  $M$ ,  $W$ ) etkileşiminin bir sonucu olarak ele alır, böylece hard problem'i bypass eder. Aynı zamanda, açık matematiksel formalizasyon sayesinde test edilebilir tahminler üretir ve yapay bilinç sistemlerinin tasarımı için somut bir yol haritası sunar.

Sonraki bölümde, FBÇT'nin formel yapısını detaylı olarak sunacağız.

## 3 Fonksiyonel Bilinç Çöküşü Teorisi

Fonksiyonel Bilinç Çöküşü Teorisi (FBÇT), bilinci zamana bağlı bir durum uzayı üzerinde işleyen, aday-üretim, çok-bileşenli değerlendirme ve olasılıksal çöküş operatörlerinden oluşan bir seçim dinamiği olarak formelleştirir. Bu bölümde teorinin temel bileşenleri, tam çöküş denklemi, değer fonksiyonu ve ajan seviyesindeki algoritmik tanım matematiksel olarak sunulmaktadır.

### 3.1 Kanonik Çöküş Fonksiyonu: $\mathcal{C}$

Bu bölümde, Fonksiyonel Bilinç Çöküşü Teorisi'nde (FBÇT) bilinçli durumun belirleyicisi olan  $\mathcal{C}$  fonksiyonunun tam, matematiksel olarak kapalı biçimini veriyoruz. Bu fonksiyon,  $(S_t, M_t, W_t)$  üçlüsünden türetilen aday bilinç durumları arasındaki rekabeti tanımlar ve "çöküş" olarak adlandırılan seçimi gerçekleştiren çekirdek mekanizmadır.

#### 3.1.1 Aday Uzayının Tanımı

Her anda sistemin erişebildiği potansiyel bilinç durumları bir aday kümesi ile temsil edilir:

$$\mathcal{X}_t = \{x_i \mid x_i \in \mathbb{X}(S_t, M_t)\}. \quad (5)$$

Bu küme, hem duyuşal durum  $S_t$  tarafından sağlanan anlık bilgileri hem de bellek  $M_t$  içindeki temsil uzaylarını kullanarak oluşturulur. Dolayısıyla  $\mathcal{X}_t$  her bilinç anında dinamik ve bağlama duyarlıdır.

Aday setinden türetilen olası bilinç durumları ise:

$$\mathcal{C}_t = \{c_i = g(x_i, S_t, M_t) \mid x_i \in \mathcal{X}_t\}, \quad (6)$$

şeklinde tanımlanır. Burada  $g$  duyuşal, bellek ve içsel temsilleri birleştiren deterministik bir haritalamadır.

#### 3.1.2 Değer Fonksiyonu: $V_i$

Her aday bilinç durumu  $c_i$ , üç kaynağa dayanan bir skor ile değerlendirilir:

$$V_i = \alpha_S f_S(c_i, S_t) + \alpha_M f_M(c_i, M_t) + \alpha_W f_W(c_i, W_t), \quad (7)$$

burada:

- $f_S$ : duyuşal uyum (sensory alignment),

- $f_M$ : bellek uyumluluğu ve çağrışımsal rezonans,
- $f_W$ : öncelikler ve değer yapısıyla uyumluluk,
- $\alpha_S, \alpha_M, \alpha_W$ : bağlama göre değişen ağırlıklandırma katsayılarıdır.

Bu üç fonksiyon, FBÇT'nin temel iddiasını yansıtır: bilinçli seçimin doğası gereği hem “şu anda gelen bilgi”, hem “geçmişte öğrenilmiş temsil”, hem de “değer/yönelim” tarafından ortaklaşa belirlendiği.

### 3.1.3 Seçim Olasılıkları

Aday durumların rekabeti softmax mekanizması ile modellenir:

$$P(c_i | S_t, M_t, W_t) = \frac{\exp(\beta V_i)}{\sum_j \exp(\beta V_j)}, \quad (8)$$

burada  $\beta > 0$  “çöküş sıcaklığı”dır.

- Yüksek  $\beta$ : daha deterministik, keskin seçim.
- Düşük  $\beta$ : daha yumuşak, dağıtık seçim.

Bu form, karar verme davranışındaki bağlamsal değişkenliği doğal olarak açıklayan bir parametre sağlar.

### 3.1.4 Çöküş Mekanizması: Kanonik Tanım

Bilinç durumu  $C_t$ , olasılık dağılımından örnekleme ile belirlenir:

$$C_t = \text{sample} (P(c_i | S_t, M_t, W_t)). \quad (9)$$

Deterministik yaklaşım kullanılmak istenirse:

$$C_t = \underset{c_i \in \mathcal{C}_t}{\text{argmax}} V_i. \quad (10)$$

### 3.1.5 Kanonik FBÇT Çöküş Operatörü

Tüm bileşenleri bir araya getirerek:

$$C_t = \mathcal{C}(S_t, M_t, W_t) = \underset{c_i \in \mathcal{X}_t}{\text{sample}} \left[ \frac{\exp(\beta [\alpha_S f_S + \alpha_M f_M + \alpha_W f_W])}{\sum_j \exp(\beta [\alpha_S f_S + \alpha_M f_M + \alpha_W f_W]_j)} \right] \quad (11)$$

Bu ifade, FBÇT'nin çekirdek mekanizmasını tek bir kapalı formda özetler. Fonksiyon:

1. duyuşal durumdan gelen anlık veriyi ( $S_t$ ),
2. bellekten gelen temsilleri ve çağrışımları ( $M_t$ ),
3. değer/öncelik sistemini ( $W_t$ ),
4. bağlama göre değişen ağırlıkları ( $\alpha_\bullet$ ),
5. seçim keskinliğini belirleyen çöküş sıcaklığını ( $\beta$ )

tek bir rekabet + çöküş süreci içerisinde birleştirir.



### 3.1.6 Çöküş Operatörünün Zorunluluğu: Neden Bu Form?

Fonksiyonel Bilinç Çöküşü Teorisi'nde (FBÇT) çöküş operatörü

$$C : (S_t, M_t, W_t) \mapsto C_t,$$

basit bir seçim mekanizması değildir; teorinin matematiksel, bilişsel ve nörobiyolojik yapısından zorunlu olarak türetilir. Bu bölümde, çöküş fonksiyonunun neden Boltzmann-softmax biçiminde olması gerektiğini ve alternatif seçim operatörlerinin neden FBÇT ile uyumsuz sonuçlar verdiğini gösteriyoruz.

**(1) Tekillik ilkesi: Bilinç tek bir içeriğe projekte olmak zorundadır.** Fenomenolojik olarak bilinç, aynı anda tek bir içeriği küresel erişime açar. Bu “global availability” ilkesi [18, 19], matematiksel olarak bir projeksiyon operatörünü gerektirir:

$$C : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{X}.$$

Dolayısıyla jeneratif uzayda  $K$  tane olası aday olsa bile, bilinç seviyesi tek bir temsile indirgenmek zorundadır.

**Değerleme fonksiyonunun zorunluluğu.** Her bilinç adımında aday durumları karşılaştırmak için sürekli, türevlenebilir, monoton bir değer fonksiyonuna ihtiyaç vardır:

$$f : X \times S \times M \times W \rightarrow \mathbb{R}.$$

Bu fonksiyon, duyuşal uygunluk ( $f_S$ ), bellek uyumluluğu ( $f_M$ ) ve motivasyonel değer ( $f_W$ ) bileşenlerinin ortak bir ölçeğe aktarılmasını sağlar. Öğrenme dinamiklerinin (Eş. 74–77) türevlenebilirlik gereksinimi,  $f$ 'nin en az  $C^1$  olmasını zorunlu kılar.

**Deterministik argmax'ın yetersizliği.**  $C_t = \arg \max f(x)$  biçimindeki deterministik bir seçim:

(i) aynı değere sahip adaylar arasında simetriyi kıramaz, (ii) duyuşal gürültü altında gözlenen mikro-zamansal fluktuasyonları açıklayamaz, (iii) öğrenme sinyallerini boğarak geri beslemeyi devre dışı bırakır.

Dolayısıyla bilinç anı deterministik olamaz; seçim *değer-duyarlı fakat stokastik* olmalıdır.

**Yönlendirilmiş rastlantısallık: Ekspansiyel familyanın zorunluluğu.** Seçim mekanizmasının şu üç özelliği aynı anda sağlaması gerekir:

1. **Değer duyarlılığı:**  $f(x)$  arttıkça seçilme olasılığı artmalıdır.
2. **Öğrenilebilirlik:** Türevlenebilir olmalı,  $\nabla f$  üzerinden güncellenebilir olmalıdır.
3. **Normalizasyon:** Olasılıkların toplamı 1 olmalıdır.

Bu üç koşulu aynı anda sağlayan *tek* dağılım ailesi, maksimum entropi prensibine göre ekspansiyel formdur:

$$\pi_t(x) \propto \exp(\beta_t f(x, S_t, M_t, W_t)).$$

Bu sonuç, Luce choice rule, Gibbs measure ve Boltzmann dağılımı arasındaki eşdeğerlikten türetilir.

**Sıcaklık parametresi  $\beta_t$ 'nin zorunluluğu.**  $\beta_t$ , belirsizlik düzeyini ve seçim keskinliğini kontrol eden bilişsel bir “gain” parametresidir:

$$\beta_t \rightarrow 0 : \text{keşif (yüksek entropi)}, \quad \beta_t \rightarrow \infty : \text{keskin, neredeyse deterministik seçim.}$$

Bu yapı, noradrenerjik ve dopaminerjik sistemlerin davranışsal kararlardaki rolüyle [20, 21] nörobiyolojik olarak doğrulanmaktadır.

**Alternatif çöküş mekanizmalarının elenmesi.** Aşağıdaki operasyonel formlar FBÇT ile uyumsuzdur:

- **Lineer seçim:**  $\pi \propto f$  ölçek değişmezliği ve normalizasyon sorunları doğurur.
- **Winner-take-all (WTA):** Türevlenebilir değildir;  $C$  öğrenmeyi durdurur ve  $\Delta M_t - \Delta W_t$  güncellemeleri çalışamaz hale gelir.
- **Tam rastgele seçim:** Değerleme tamamen ihmal edilir; bilinç deneyimi ile davranış arasında gözlenen korelasyon açıklanamaz.
- **Deterministik argmax:** Mikro-varyasyonlar kaybolur, soft attention mekanizmaları modelden çıkmış olur.

Bu nedenle üç temel koşulu (değer duyarlılığı, türevlenebilirlik, normalizasyon) sağlayan *tek* çöküş mekanizması Boltzmann-softmax'tır.

**FBÇT içindeki çöküşün işlevsel rolü.** Çöküş yalnızca o anki bilinç içeriğini üretmez; aynı zamanda gelecekteki güncellemelerin kaynağıdır:

$$(M_{t+1}, W_{t+1}) = F(M_t, W_t, S_t, C_t).$$

Bu geri-besleme döngüsü sayesinde:

- her seçim bellek izleri bırakır (path dependence),
- öncelikler yeniden şekillenir,
- uzun ölçekli kimlik ve kişilik sürekliliği ortaya çıkar,
- $\alpha$  ağırlıkları bağlamsal olarak yeniden düzenlenir.

**Sonuç: Boltzmann-softmax formu matematiksel bir zorunluluktur.** FBÇT'nin yapısal ilkeleri, (i) teklik, (ii) değer-duyarlı stokastik seçim, (iii) enerji tabanlı tutarlılık, (iv) öğrenilebilirlik, (v) nörobiyolojik uyumluluk

koşullarını aynı anda sağlar. Bu koşulları aynı çerçevede karşılayan *tek* form:

$$C_t = \text{sample} \left[ \frac{\exp(\beta_t f(x))}{\sum_j \exp(\beta_t f(x_j))} \right]$$

şeklindeki Boltzmann-softmax operatörüdür.

Dolayısıyla çöküş fonksiyonu, FBÇT'de keyfi bir seçim değil, teoremin yapısından zorunlu olarak türetilen tek matematiksel formdur.

Çöküş operatörünün ölçü-teorik ve sürekli uzaylara genişletilmiş formu Ek A'da verilmiştir.

### 3.2 Bellek ve Öncelik Sistemlerinin İç Yapısı

FBÇT'nin temel bileşenleri olan bellek yapısı  $M_t$  ve öncelik sistemi  $W_t$ , tekil skalar değerler değil, zengin iç yapıya sahip çok-bileşenli sistemlerdir. Bu bölümde her iki yapının matematiksel organizasyonu, alt-bileşenleri ve değer fonksiyonuna entegrasyonu detaylandırılmaktadır.

### 3.2.1 Bellek Yapısı: $M_t$

Bellek, dört etkileşimli alt-sistem olarak modellenir:

$$M_t = (M_t^{\text{sem}}, M_t^{\text{epi}}, M_t^{\text{assoc}}, M_t^{\text{aff}}), \quad (12)$$

burada her bileşen farklı bir bellek türünü temsil eder.

**(a) Semantik Bellek:**  $M_t^{\text{sem}}$  Uzun süreli, soyut bilgi yapılarını içerir. Graf-temelli bir manifold olarak modellenir:

$$M_t^{\text{sem}} = (\mathcal{C}_t, E_t^{\text{sem}}, \mathbf{h}^{\text{sem}}), \quad (13)$$

burada:

- $\mathcal{C}_t$ : kavram düğümleri kümesi (örn. "ev", "anne", "tehlike"),
- $E_t^{\text{sem}} \subseteq \mathcal{C}_t \times \mathcal{C}_t$ : semantik ilişkiler (örn. "köpek"-"hayvan" hiyerarşisi),
- $\mathbf{h}^{\text{sem}} : \mathcal{C}_t \rightarrow \mathbb{R}^{d_{\text{sem}}}$ : kavram embedding fonksiyonu.

Bir aday durum  $x$ 'in semantik bellekle uyumu:

$$s^{\text{sem}}(x, M_t^{\text{sem}}) = \max_{c \in \mathcal{C}_t} \exp\left(-\frac{\|\mathbf{h}_x - \mathbf{h}^{\text{sem}}(c)\|^2}{\sigma_{\text{sem}}^2}\right), \quad (14)$$

burada  $\mathbf{h}_x \in \mathbb{R}^{d_{\text{sem}}}$  aday durumun semantik temsilidir.

**Nörobiyolojik karşılık:** Temporal korteks, inferior frontal gyrus.

**(b) Episodik Bellek:**  $M_t^{\text{epi}}$  Kişisel, zaman damgalı deneyimlerin kayıdır:

$$M_t^{\text{epi}} = \{(S_\tau, C_\tau, \tau) \mid \tau \in [t - T_{\text{memory}}, t]\}, \quad (15)$$

burada  $T_{\text{memory}}$  episodik bellek penceresinin uzunluğudur.

Bir aday durumun episodik bellekle uyumu, geçmiş deneyimlere olan benzerlik üzerinden hesaplanır:

$$s^{\text{epi}}(x, S_t, M_t^{\text{epi}}) = \sum_{\tau \in M_t^{\text{epi}}} \underbrace{\exp(-\lambda_{\text{time}}(t - \tau))}_{\text{zaman çürümesi}} \cdot \underbrace{\exp\left(-\frac{\|(x, S_t) - (C_\tau, S_\tau)\|^2}{\sigma_{\text{epi}}^2}\right)}_{\text{benzerlik}}, \quad (16)$$

burada  $\lambda_{\text{time}} > 0$  zaman çürüme hızıdır (yakın geçmiş daha etkili).

**Nörobiyolojik karşılık:** Hipokampus.

**(c) Çağrışımsal Ağ:**  $M_t^{\text{assoc}}$  Kavramlar arası ilişkilendirmeleri kodlar. Sparse bir ağırlık matrisi ile temsil edilir:

$$M_t^{\text{assoc}} = A_t, \quad A_{ij} \in [0, 1], \quad (17)$$

burada  $A_{ij}$ , kavram  $i$  ile kavram  $j$  arasındaki çağrışım gücüdür.

**Sparse yapı:** Pratikte  $A_t$  yalnızca aktif bağlantılar için değer taşır:

$$A_{ij} \neq 0 \iff (i, j) \in E_t^{\text{assoc}}, \quad (18)$$

bu sayede  $n$  kavram için  $\mathcal{O}(n^2)$  yerine  $\mathcal{O}(|E_t^{\text{assoc}}|)$  bellek kullanılır.

Çağrışımsal aktivasyon:

$$s^{\text{assoc}}(x, M_t^{\text{assoc}}) = \sum_{i, j \in \text{active}(x)} A_{ij}, \quad (19)$$

burada  $\text{active}(x)$ , aday durum  $x$  tarafından aktive edilen kavram indeksleridir.

**Nörobiyolojik karşılık:** Prefrontal-temporal bağlantılar, default mode network.

(d) **Duygusal Bellek:**  $M_t^{\text{aff}}$  Her kavramın duygusal yükünü kodlar:

$$M_t^{\text{aff}} = \{(c, v_c, a_c) \mid c \in \mathcal{C}_t\}, \quad (20)$$

burada:

- $v_c \in [-1, +1]$ : valans (pozitif/negatif his),
- $a_c \in [0, 1]$ : arousal (uyarılmışlık).

Duygusal skor:

$$s^{\text{aff}}(x, M_t^{\text{aff}}) = \sum_{c \in \text{active}(x)} (\omega_v v_c + \omega_a a_c), \quad (21)$$

burada  $\omega_v, \omega_a$  ağırlık parametreleridir.

**Nörobiyolojik karşılık:** Amygdala, orbitofrontal cortex.

**Bellek Bileşenlerinin Entegrasyonu.** Dört bellek skoru, ağırlıklı toplam ile birleştirilir:

$$s_M(x, S_t, M_t) = \beta_{\text{sem}} s^{\text{sem}} + \beta_{\text{epi}} s^{\text{epi}} + \beta_{\text{assoc}} s^{\text{assoc}} + \beta_{\text{aff}} s^{\text{aff}}, \quad (22)$$

burada  $\beta_{\text{sem}}, \beta_{\text{epi}}, \beta_{\text{assoc}}, \beta_{\text{aff}} \geq 0$  her bellek türünün göreceli önemini belirler.

Bu toplam skor, değer fonksiyonuna şu şekilde dahil edilir:

$$U_j^{\text{modulated}}(x, S_t, M_t) = U_j^{\text{base}}(x, S_t) \cdot (1 + \alpha_j s_M(x, S_t, M_t)), \quad (23)$$

burada  $U_j^{\text{base}}$  temel değer bileşeni (3.4) ve  $\alpha_j \geq 0$  modülasyon katsayısıdır.

### 3.2.2 Öncelik Sistemi: $W_t$

Öncelik sistemi, farklı değer bileşenlerine verilen göreceli önemleri kodlar:

$$W_t = (W_t^{\text{homeo}}, W_t^{\text{reward}}, W_t^{\text{threat}}, W_t^{\text{goal}}, W_t^{\text{social}}) \in \Delta^{k-1}, \quad (24)$$

burada  $\Delta^{k-1}$  simplex'tir ( $\sum_i W_t^{(i)} = 1$ ).

(a) **Homeostatik Öncelikler:**  $W_t^{\text{homeo}}$  Temel fizyolojik gereksinimlerin (açlık, susuzluk, uyku, ağrı) önceliğini belirler:

$$W_t^{\text{homeo}} = \phi_{\text{homeo}}(H_t), \quad (25)$$

burada  $H_t$  interoceptive durum vektörüdür (kalp atışı, kan şekeri, uyku borcu, vb.) ve  $\phi_{\text{homeo}}$  bu durumu öncelik değerine çeviren bir fonksiyondur.

**Tipik form:**

$$W_t^{\text{homeo}} = \sigma(\mathbf{w}_{\text{homeo}}^\top H_t), \quad (26)$$

burada  $\mathbf{w}_{\text{homeo}}$  öğrenilebilir ağırlıklar ve  $\sigma$  sigmoid fonksiyonudur.

**Nörobiyolojik karşılık:** Hipotalamus, insula.

(b) **Ödül Öncelikleri:**  $W_t^{\text{reward}}$  Dopaminerjik ödül sisteminin temsilcidir:

$$W_t^{\text{reward}} \in [0, 1]^{k_r}, \quad (27)$$

burada  $k_r$  farklı ödül kategorilerinin sayısıdır (örn. yiyecek, sosyal onay, parasal kazanç).

Bu bileşen, deneyim yoluyla öğrenilir (aşağıda detaylandırılacak).

**Nörobiyolojik karşılık:** Ventral tegmental area (VTA), nucleus accumbens, ventral striatum.

(c) **Tehdit Öncelikleri:**  $W_t^{\text{threat}}$  Tehlike algısı ve savunma davranışlarının önceliğini kodlar:

$$W_t^{\text{threat}} \in [0, 1]^{k_t}, \quad (28)$$

Yüksek  $W_t^{\text{threat}}$ , kaçınma ve savunma davranışlarının ağır basmasına yol açar.

**Nörobiyolojik karşılık:** Amygdala, periaqueductal gray.

(d) **Hedef Öncelikleri:**  $W_t^{\text{goal}}$  Uzun vadeli görevler ve planlar için öncelik:

$$W_t^{\text{goal}} \in [0, 1]^{k_g}, \quad (29)$$

Bu bileşen, meta-bilişsel kontrol ve yönetici işlevlerle ilişkilidir.

**Nörobiyolojik karşılık:** Dorsolateral prefrontal cortex.

(e) **Sosyal Öncelikler:**  $W_t^{\text{social}}$  Sosyal ipuçları ve normların önemi:

$$W_t^{\text{social}} \in [0, 1]^{k_s}, \quad (30)$$

**Nörobiyolojik karşılık:** Temporoparietal junction, medial prefrontal cortex.

**Öncelik Sisteminin Değer Fonksiyonuna Entegrasyonu.** 3.4'te tanımlanan değer fonksiyonu  $U(x, S_t, M_t) \in \mathbb{R}^k$ , bu önceliklerle şu şekilde ağırlıklandırılır:

$$f(x, S_t, M_t, W_t) = \langle W_t, U^{\text{modulated}}(x, S_t, M_t) \rangle = \sum_{j=1}^k W_t^{(j)} U_j^{\text{modulated}}(x, S_t, M_t), \quad (31)$$

burada  $U_j^{\text{modulated}}$  (23) ile tanımlanan bellek-modüle edilmiş değer bileşenidir.

### 3.2.3 Güncelleme Dinamikleri

Her çöküş  $C_t$  sonrasında, bellek ve öncelik sistemleri güncellenir:

$$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t, R_t), \quad (32)$$

burada  $R_t$  içsel veya dışsal ödül/ceza sinyalidir.

**Bellek Bileşenlerinin Güncellemesi. Semantik bellek:** Embedding'ler yavaş öğrenme ile güncellenir:

$$\mathbf{h}_{t+1}^{\text{sem}}(c) = \mathbf{h}_t^{\text{sem}}(c) + \eta_{\text{sem}} \nabla_h \mathcal{L}_{\text{context}}(c, C_t, S_t), \quad (33)$$

burada  $\mathcal{L}_{\text{context}}$  bağlamsal uygunluk kaybı (örn. contrastive learning) ve  $\eta_{\text{sem}} \ll 1$ .

**Episodik bellek:** Yeni deneyim eklenir:

$$M_{t+1}^{\text{epi}} = M_t^{\text{epi}} \cup \{(S_t, C_t, t)\}. \quad (34)$$

Bellek kapasitesi aşıldığında, en eski veya en az erişilen kayıtlar silinir (örn. FIFO veya relevance-based pruning).

**Çağrışımsal ağ:** Hebbian-benzeri güncelleme:

$$\Delta A_{ij} = \eta_{\text{assoc}} a_i(C_t) a_j(C_t) - \gamma_{\text{assoc}} A_{ij}, \quad (35)$$

burada  $a_i(C_t)$  kavram  $i$ 'nin  $C_t$ 'deki aktivasyon düzeyi,  $\gamma_{\text{assoc}}$  unutmaya oranıdır.

**Duygusal bellek:** Valans ve arousal etiketleri yumuşak güncelleme ile:

$$v_{c,t+1} = (1 - \eta_v) v_{c,t} + \eta_v \phi_{\text{val}}(C_t, S_t), \quad (36)$$

$$a_{c,t+1} = (1 - \eta_a) a_{c,t} + \eta_a \phi_{\text{ar}}(C_t, S_t), \quad (37)$$

burada  $\phi_{\text{val}}$  ve  $\phi_{\text{ar}}$ , 3.11.1'te tanımlanan valans ve arousal fonksiyonlarıdır.

Böylece  $s_M(x, S_t, M_t)$ , bellek içeriğini doğrudan değer fonksiyonuna (23) üzerinden aktarır; bu modüle edilmiş değerler, çöküş mekanizması  $\mathcal{C}$  içinde kullanılan skor fonksiyonuna (3.5) ile beslenir.

**Öncelik Bileşenlerinin Güncellemesi.** **Homeostatik öncelikler:** Beden durumu değişimine göre otomatik güncellenir:

$$W_{t+1}^{\text{homeo}} = \phi_{\text{homeo}}(H_{t+1}). \quad (38)$$

**Ödül öncelikleri:** Temporal difference (TD) learning ile:

$$W_{t+1}^{\text{reward}} = W_t^{\text{reward}} + \alpha_r \delta_t \nabla_W f(C_t, S_t, M_t, W_t), \quad (39)$$

burada  $\delta_t = R_t + \gamma V(S_{t+1}) - V(S_t)$  TD hatasıdır.

**Tehdit öncelikleri:** Tehdit deneyimleriyle hassaslaşır:

$$W_{t+1}^{\text{threat}} = W_t^{\text{threat}} + \alpha_t \mathbb{I}_{\text{threat}}(S_t, C_t, R_t), \quad (40)$$

burada  $\mathbb{I}_{\text{threat}}$  tehdit sinyali göstergesidir ( $R_t < 0$  ve tehdit içeriği varsa 1).

**Hedef öncelikleri:** Meta-öğrenme veya görev başarısına göre:

$$W_{t+1}^{\text{goal}} = W_t^{\text{goal}} + \alpha_g \nabla_W J_{\text{task}}(C_t, S_t), \quad (41)$$

burada  $J_{\text{task}}$  görev başarımlı fonksiyondur.

**Sosyal öncelikler:** Sosyal geri bildirimine göre:

$$W_{t+1}^{\text{social}} = W_t^{\text{social}} + \alpha_s F_{\text{social}}(S_t, C_t, R_t), \quad (42)$$

burada  $F_{\text{social}}$  sosyal onay/ret sinyalini işleyen fonksiyondur.

**Simplex kısıtı:** Her güncellemeden sonra, öncelikler normalize edilir:

$$W_{t+1} \leftarrow \frac{W_{t+1}}{\sum_j W_{t+1}^{(j)}}. \quad (43)$$

Özetle,

$$\boxed{\begin{aligned} s_M(x, S_t, M_t) &\rightarrow U^{\text{modulated}}(x, S_t, M_t) \\ U^{\text{modulated}}(x, S_t, M_t), W_t &\rightarrow f(x, S_t, M_t, W_t) \\ f &\rightarrow \pi_t \rightarrow C_t \\ C_t, R_t &\rightarrow (M_{t+1}, W_{t+1}) \end{aligned}} \quad (44)$$

### 3.2.4 Nörobiyolojik Eşlemeler: Özet Tablo

Table 2: Bellek ve öncelik bileşenlerinin nörobiyolojik karşılıkları

Bileşen	FBÇT Notasyonu	Nöral Substrat
<i>Bellek Sistemi (<math>M_t</math>)</i>		
Semantik	$M_t^{\text{sem}}$	Temporal korteks, IFG
Episodik	$M_t^{\text{epi}}$	Hipokampus, MTL
Çağrışımsal	$M_t^{\text{assoc}}$	PFC-temporal bağlantılar, DMN
Duygusal	$M_t^{\text{aff}}$	Amygdala, OFC
<i>Öncelik Sistemi (<math>W_t</math>)</i>		
Homeostatik	$W_t^{\text{homeo}}$	Hipotalamus, insula
Ödül	$W_t^{\text{reward}}$	VTA, NAcc, ventral striatum
Tehdit	$W_t^{\text{threat}}$	Amygdala, PAG
Hedef	$W_t^{\text{goal}}$	dIPFC
Sosyal	$W_t^{\text{social}}$	TPJ, mPFC

**Kısaltmalar:** IFG = inferior frontal gyrus, MTL = medial temporal lobe, PFC = prefrontal cortex, DMN = default mode network, OFC = orbitofrontal cortex, VTA = ventral tegmental area, NAcc = nucleus accumbens, PAG = periaqueductal gray, dIPFC = dorsolateral PFC, TPJ = temporoparietal junction, mPFC = medial PFC.

### 3.2.5 Sonuç

Bu detaylı yapılandırma, FBÇT'nin  $M$  ve  $W$  bileşenlerinin:

1. **Zengin iç yapıya** sahip olduğunu (her biri birden fazla alt-sistem),
2. **Nörobiyolojik olarak haklı** eşlemeler taşıdığını,
3. **Değer fonksiyonuna açıkça entegre** edildiğini ((23), (3.5)),
4. **Deneyimle öğrenilebilir ve güncellenebilir** dinamiklere sahip olduğunu

göstermektedir. Bu yapı, teorinin hem nörobiliml açıdan gerçekçi hem de hesaplamalı olarak uygulanabilir olmasını sağlar.

### 3.3 Durum Uzayları ve Temel Değişkenler

FBÇT, her ayrık zaman adımında  $t \in \mathbb{Z}$ , sistemin bilişsel durumunu dört ana vektörle temsil eder: duyuşsal durum  $S_t$ , bellek/yorumlama durumu  $M_t$ , öncelik/ağırlık vektörü  $W_t$  ve iç temsil durumu  $X_t$ .

**Duyuşsal durum.**

$$S_t \in \mathbb{R}^{n_S}. \quad (45)$$

$S_t$ , o anda işlenmiş tüm duyuşsal girdilerin (basit özelliklerden yüksek seviye algısal temsillere kadar) birleşik temsildir. Pratikte  $S_t$ , farklı modaliteler için ayrılmış alt-uzaylardan oluşabilir:

$$S_t = (s_t^{\text{vis}}, s_t^{\text{aud}}, s_t^{\text{som}}, s_t^{\text{int}}, \dots), \quad (46)$$

burada her bileşen ilgili modalitenin işlenmiş çıktısını temsil eder.

**Bellek/yorumlama durumu.**

$$M_t \in \mathbb{R}^{n_M}. \quad (47)$$

$M_t$ , episodik ve semantik bellek, öğrenilmiş şemalar ve duygusal etiketler dâhil olmak üzere geçmiş deneyimden türetilen yapısal bilgiyi temsil eder. Kavramsal olarak

$$M_t = (M_t^e, M_t^s, M_t^a) \quad (48)$$

şeklinde (episodik, semantik, affektif) alt bileşenlere ayrılabilir; ancak teorik çekirdekte tek bir durum vektörü olarak ele alınır.

**Öncelik/ağırlık vektörü.**

$$W_t \in \Delta^{k-1}, \quad (49)$$

burada

$$\Delta^{k-1} = \left\{ w \in \mathbb{R}^k \mid w_i \geq 0, \sum_{i=1}^k w_i = 1 \right\} \quad (50)$$

$k$  boyutlu olasılık simpleksidir.  $W_t$ , sistemin farklı değer bileşenlerine (ör. *güvenlik*, *ödül*, *sosyal değer*, *epistemik değer*, *benlik tutarlılığı*) verdiği bağıl öncelikleri kodlar ve hem nispeten sabit (doğuştan, kültürel) hem de öğrenilmiş bileşenler içerir:

$$W_t = W_t^{\text{innate}} + W_t^{\text{learned}}. \quad (51)$$

**İç temsil durumu.**

$$X_t \in \mathbb{R}^{n_x}, \quad (52)$$

$X_t$ , sistemin o anda erişilebilir iç temsil uzayını ifade eder. Kavramsal olarak iki alt bileşene ayrılır:

$$X_t = (x_t^{\text{world}}, x_t^{\text{self}}), \quad x_t^{\text{world}} \in \mathbb{R}^{d_w}, \quad x_t^{\text{self}} \in \mathbb{R}^{d_s}, \quad (53)$$

burada  $x_t^{\text{world}}$  dış dünyaya ilişkin temsilleri,  $x_t^{\text{self}}$  ise sistemin kendi durumuna ilişkin temsilleri içermektedir.

**Fonksiyonel durum uzayı.** Tüm bileşenler birlikte bir fonksiyonel durum uzayı tanımlar:

$$\Omega = \mathcal{S} \times \mathcal{M} \times \mathcal{W} \subseteq \mathbb{R}^{n_s} \times \mathbb{R}^{n_M} \times \Delta^{k-1}, \quad (54)$$

ve bilinç dinamiği,  $\Omega$  üzerinde tanımlı bir işlem ailesi olarak modellenir. İç temsil  $X_t$ , çöküş işleminin çıktısı olarak ele alınır; dolayısıyla  $\Omega$  ile  $X$  arasındaki köprü, çöküş mekanizmasıdır.

**Enerji-bazlı yorum.** Değer fonksiyonu bir enerji fonksiyonunun negatifi olarak yorumlanabilir:

$$E(x) = -f(x, S_t, M_t, W_t), \quad (55)$$

ve politika dağılımı Boltzmann dağılımı formunu alır:

$$\pi_t(x) \propto \exp(-\beta_t E(x)). \quad (56)$$

Yüksek  $\beta_t$  limitinde ( $\beta_t \rightarrow \infty$ ), çöküş deterministik enerji minimizasyonuna yakınsar:

$$C_t \approx \arg \min_x E(x). \quad (57)$$

Bu yorum, FBÇT'yi statistical mechanics, free-energy principle ve enerji-tabanlı modeller (örn. Hopfield networks, restricted Boltzmann machines) ile ilişkilendirir.

### 3.4 Aday Üretimi ve Değer Fonksiyonu

FBÇT'de bilinçli durum, tek adımlı deterministik bir fonksiyon uygulamasının sonucu değil, bir *aday üretim süreci* ve bu adaylar üzerinde tanımlı bir *çok bileşenli değer fonksiyonu* üzerinden hesaplanan *olasılıksal çöküş* işlemiyle belirlenir.

#### 3.4.1 Jeneratif Aday Üretici: $G$

FBÇT'de çöküş operatörü  $C_t$  tek başına çalışmaz; her zaman adımında önce, iç temsil uzayı  $X$  üzerinde olası bilişsel durumları içeren bir aday kümesi  $X_t$  üretilir. Bu küme, duysal durum  $S_t$ , bellek durumu  $M_t$  ve öncelik vektörü  $W_t$  tarafından koşullanan bir jeneratif çekirdek ile tanımlanır. Bu alt bölümde  $G$  operatörünün matematiksel formunu, hesaplamalı gerçekleşmesini ve nörobiyolojik karşılıklarını sistematik biçimde veriyoruz.

#### 3.4.2 Jeneratif çekirdeğin tanımı

Her zaman adımı için,  $(S_t, M_t, W_t)$  üçlüsüne koşullu bir olasılık çekirdeği tanımlarız:

$$\mu_t(dx) = G(S_t, M_t, W_t)(dx), \quad x \in X.$$

Operasyonel olarak  $G$ , iç temsil uzayı üzerinde  $K$  adet aday örnek üreten bir süreçtir:

$$x_t^{(k)} \sim \mu_t(dx) \quad (k = 1, \dots, K), \quad X_t = \{x_t^{(1)}, \dots, x_t^{(K)}\}.$$

Dolayısıyla  $G$ , bilinç çöküşünün gerçekleşeceği aday uzayı  $X_t$ 'yi tanımlar;  $C_t$  ise bu uzay üzerinde seçim yapan operatördür (bkz. Bölüm 3.7).



### 3.4.3 Enerji-tabanlı birleşik form

$G$  operatörünün, teoremin geri kalanıyla uyumlu olması için hem *öğrenilebilir* (gradyanları tanımlı), hem de *değer duyarlı* olması gerekir. Bu nedenle  $G$ , koşullu enerji-tabanlı bir model olarak tanımlanır:

$$p(x \mid S_t, M_t, W_t) = \frac{1}{Z_t} \exp(-E(x; S_t, M_t) + V(x; W_t)),$$

burada  $Z_t$  normalizasyon sabitidir:

$$Z_t = \int_X \exp(-E(x; S_t, M_t) + V(x; W_t)) dx.$$

Buradaki  $E$  terimi duyuşal ve bellek uygunluğunu,  $V$  terimi ise değer/öncelik yönelimini temsil eder.

### 3.4.4 Enerji bileşenleri ve değer potansiyeli

**Duyuşal enerji terimi.** Duyuşal durum  $S_t$ , işlenmiş duyuşal temsilleri iç temsil uzayına projeksiyon yapan bir kodlayıcı ile modellenir:

$$\phi_S : S \rightarrow X.$$

Duyuşal uygunluk,  $x$ 'in bu projeksiyona yakınlığı ile ölçülür:

$$E_S(x; S_t) = \frac{1}{2\sigma_S^2} \|x - \phi_S(S_t)\|^2.$$

**Bellek enerji terimi.** Bellek durumu  $M_t$ , Bölüm 3.2'de tanımlanan episodik ve semantik alt sistemleri içerir. Bu yapının jeneratif düzeydeki etkisini,  $M_t$ 'den örneklenen temsillerin ortalama enerjisi ile ifade ederiz. Bellek kodlayıcı

$$\phi_M : \mathcal{M} \rightarrow X$$

ile

$$E_M(x; M_t) = \mathbb{E}_{m \sim M_t} \left[ \frac{1}{2\sigma_M^2} \|x - \phi_M(m)\|^2 \right].$$

Pratikte bu beklenti,  $M_t$ 'deki son deneyimlerin sonlu bir örnekleme üzerinden yaklaşık olarak hesaplanır.

**Toplam enerji.** Duyuşal ve bellek katkıları birleştirilerek toplam enerji elde edilir:

$$E(x; S_t, M_t) = E_S(x; S_t) + \lambda_M E_M(x; M_t),$$

burada  $\lambda_M \geq 0$  bellek katkısının göreceli ağırlığını belirler.

**Değer potansiyeli.** Öncelik vektörü  $W_t \in \Delta_{k-1}$ , hangi aday temsillerin işlevsel olarak daha avantajlı olduğunu kodlar. Aday durumun değer-duyarlı özelliklerini çıkaran bir fonksiyon

$$\psi : X \rightarrow \mathbb{R}^k$$

ile değer potansiyelini

$$V(x; W_t) = \lambda_W W_t^\top \psi(x),$$

şeklinde tanımlarız; burada  $\lambda_W \geq 0$ , değer etkisinin gücünü belirler.

Bu bileşenler birleştirildiğinde  $G$  tarafından tanımlanan koşullu dağılımın nihai biçimi:

$$p(x \mid S_t, M_t, W_t) = \frac{1}{Z_t} \exp \left( -\frac{\|x - \phi_S(S_t)\|^2}{2\sigma_S^2} - \lambda_M \mathbb{E}_{m \sim M_t} \frac{\|x - \phi_M(m)\|^2}{2\sigma_M^2} + \lambda_W W_t^\top \psi(x) \right).$$

### 3.4.5 Somut örnek

Jeneratif sürecin işleyişini somutlaştırmak için şu senaryoyu ele alalım: Bir kişi, gece karanlığında evin bahçesinde belirsiz bir hareket algılıyor.

**Durum:**

- $S_t$ : Düşük çözünürlüklü, gölgesel bir hareket (belirsiz görsel giriş),
- $M_t$ : Geçmiş deneyimler (“kedimiz bahçede dolaşır”, “komşu geç saatte gelir”, “mahallede hırsızlık olmuştu”),
- $W_t$ : Yüksek tehdit önceliği (gece, yalnızlık, genel güvensizlik).

**Jeneratif süreç:**

- Duyusal enerji**  $E_S$ , kabaca “hareket eden bir nesne” şeklinde bir temsil  $\phi_S(S_t)$  üretir; ayrıntı düşüktür, sadece silüet bilgisi taşır.
- Bellek enerjisi**  $E_M$ ,  $M_t$ ’deki benzer durumları çağırır:

- $m_1$  : “geçen sene bahçede kedimiz vardı”,
- $m_2$  : “komşu bazen geç saatte gelir”,
- $m_3$  : “mahallede kısa süre önce hırsızlık yaşanmıştı”.

Bu kayıtlardan  $\phi_M(m_i)$  temsilleri çıkarılır ve olası aday  $x$  ile olan uzaklıkları toplam enerjiye katkı verir.

- Değer potansiyeli**  $V$ ,  $W_t$ ’deki tehdit bileşeninin yüksek olması nedeniyle, tehdit içeren yorumlar ( $x_{\text{hırsız}}$ ) için daha yüksek potansiyel üretir:

$$V(x_{\text{hırsız}}; W_t) > V(x_{\text{kedi}}; W_t).$$

**Adayların oluşumu.** Bu kombinasyon sonucunda  $G$  operatörü, yaklaşık olarak şu tür adaylar üretebilir:

$$\begin{aligned} X_t = \{ & x^{(1)} : \text{“hırsız”} \quad (\text{yüksek } V, \text{ orta } E_M), \\ & x^{(2)} : \text{“komşu”} \quad (\text{orta } V, \text{ düşük } E_M), \\ & x^{(3)} : \text{“kedi”} \quad (\text{daha düşük } V, \text{ düşük } E_M), \\ & x^{(4)} : \text{“belirsiz gölge”} \quad (\text{çok düşük } V, \text{ düşük } E_S) \}. \end{aligned}$$

Sonraki adımda çöküş operatörü  $C_t$  (Bölüm 3.7), bu adaylar arasından değer fonksiyonu  $f$  ve sıcaklık parametresi  $\beta_t$ ’ye bağlı olarak seçim yapar. Yüksek tehdit durumunda (yüksek etkinlik gösteren ilgili bileşenler), “hırsız” yorumuna karşılık gelen durumun seçilme olasılığı artar; tehdit bileşeni daha düşük bir bağlamda ise “kedi” veya “komşu” gibi benign yorumlar ön plana çıkar. Böylece  $G$ , duyusal belirsizlik, bellek bağlamı ve motivasyonel durum arasındaki etkileşimi açık biçimde modellemiş olur.

### 3.4.6 Parametre belirleme

Jeneratif sürecin parametreleri

$$\theta_G = \{\sigma_S, \sigma_M, \lambda_M, \lambda_W\}$$

iki ana yaklaşımla belirlenebilir:

**Yaklaşım 1: Sabit hiperparametreler.** Parametreler sabit hiperparametreler olarak seçilir ve elle ayarlama veya çapraz doğrulama ile optimize edilir. Bu yaklaşım basit bir implementasyon sağlar ancak bağlama duyarlı adaptasyon sınırlıdır.

**Yaklaşım 2: Öğrenilebilir parametreler (önerilen).** FBÇT, jeneratif parametrelerin deneyim üzerinden güncellenmesini önerir. Parametreler, jeneratif kaliteyi ve davranışsal başarıyı yansıtan bir kayıp fonksiyonu ile öğrenilir:

$$\theta_{G,t+1} = \theta_{G,t} + \alpha_G \nabla_{\theta} \mathcal{L}_G(\theta_{G,t}),$$

burada  $\alpha_G$  öğrenme oranı,  $\mathcal{L}_G$  ise jeneratif kalite kayıp fonksiyonudur (örneğin yeniden-oluşturma hatası, VAE kaybı veya davranışsal ödül  $R_t$  ile ağırlıklandırılmış bir loss). Böylece  $G$ , farklı bağlamlarda (tehdit vs. güvenlik, aşinalık vs. yenilik) farklı enerji dengelerine uyum sağlayabilir. Detaylı güncelleme kuralları Bölüm 3.2.3’de verilmiştir.

### 3.4.7 Hesaplamalı gerçekleştirme

$p(x | S_t, M_t, W_t)$ ’den doğrudan örnekleme, normalizasyon sabiti  $Z_t$ ’nin hesaplanmasını gerektirir. Genel durumda bu hesap intractable olduğundan, pratikte iki sınıf yaklaşım kullanılır.

**Yaklaşım A: Langevin dinamiği ile örnekleme.** Enerji gradyanına dayalı bir MCMC süreci:

$$x_{i+1} = x_i - \epsilon \nabla_x E(x_i; S_t, M_t) + \epsilon \nabla_x V(x_i; W_t) + \sqrt{2\epsilon} \eta_i, \quad (58)$$

burada  $\eta_i \sim \mathcal{N}(0, I)$  ve  $i$  iterasyon indeksidir. Yeterince uzun bir süre çalıştırıldığında bu süreç, teorik olarak hedef dağılıma yakınsamayı garanti eder; ancak hesaplama maliyeti yüksek olabilir.

**Yaklaşım B: Amortize çıkarım (önerilen).** Daha pratik bir yaklaşımda, parametrik bir dağılım  $q_{\phi}(x | S, M, W)$ ,  $p(x | S, M, W)$ ’yi yaklaşık olarak öğrenir:

$$\mathcal{L}_{\text{amortize}}(\phi) = \mathbb{E}_{S,M,W} [D_{\text{KL}}(q_{\phi}(\cdot | S, M, W) \| p(\cdot | S, M, W))]. \quad (59)$$

$q_{\phi}$  bir kez eğitildikten sonra, her zaman adımında örnekleme, tek bir ileri geçiş ile gerçekleştirilir:

$$x_t^{(k)} \sim q_{\phi}(\cdot | S_t, M_t, W_t), \quad X_t = \{x_t^{(k)}\}_{k=1}^K.$$

Bölüm 1’te sunulan ajan mimarisinde bu ikinci yaklaşım kullanılmaktadır.

### 3.4.8 Nörobiyolojik substrat

Jeneratif operatör  $G$ ’nin bileşenleri, beynin farklı sistemlerine doğal olarak eşlenebilir. Tablo 3, modeldeki soyut fonksiyonların olası nöral karşılıklarını özetler.

Bileşen	Olası nöral substrat	İşlev
$\phi_S(S_t)$	V1-V4, birincil/sekonder duyuşal korteks	Duyuşal encoding ve erken temsil
$\phi_M(m)$	Hippocampus, medial temporal lob	Episodik bellek çağırma, pattern completion
$\psi(x)$	OFC, vmPFC	Değer boyutlu özellik çıkarımı
$E_S + E_M$	Temporal-parietal ağ	Uygunluk/benzerlik hesabı, bağlamsal entegrasyon
$V(x; W_t)$	VTA, NAcc (dopaminerjik devreler)	Motivasyonel ağırlıklandırma, ödül/tehdit biası
Örnekleme kontrolü	DLPFC, ACC	Jeneratif sürecin kontrolü, hipotez seçimi

Table 3: Jeneratif operatör  $G$ ’nin bileşenleri için olası nörobiyolojik eşlemeler.

Bu eşleme, Bölüm 6.2’te formüle edilen test edilebilir öngörülerin temelini oluşturur. Özellikle  $G$ ’nin ürettiği adayların çeşitliliği ile prefrontal aktivite arasındaki ilişki (Tahmin 3) doğrudan bu yapıdan türetilmektedir.

### 3.4.9 Özet ve teorik rolü

Özetle  $G$ , FBÇT içinde üç düzeyde merkezi bir rol oynar: (i) matematiksel düzeyde,  $C_t$  ile aynı eksponansiyel familya içinde yer alan enerji tabanlı bir jeneratif operatör olarak; (ii) hesaplamalı düzeyde, duyusal belirsizliği, bellek bağlamını ve motivasyonel ağırlıkları içsel aday temsillere dönüştüren bir hipotez üretim mekanizması olarak; (iii) nörobiyolojik düzeyde ise, hippocampal yeniden-oynatım, prefrontal jeneratif kontrol ve değer sistemleri arasındaki etkileşimi yakalamaya yönelik birleşik bir şema olarak.

Bilinç çöküşü  $C_t$ 'nin seçtiği içerikler, bu jeneratif uzay üzerinden belirlenir. Dolayısıyla  $G$ 'nin yapısı, yalnızca teknik bir detay değil, teorinin fenomenolojik zenginlik, esneklik ve bağlama duyarlılık iddialarının doğrudan taşıyıcısıdır.

### 3.5 Fonksiyonel Değer Fonksiyonu: $f$

**Bu bölümün kapsamı ve bölüm 3.2 ile ilişkisi.** Bölüm 3.2'de bellek yapısı  $M_t$  ve öncelik sistemi  $W_t$ 'nin iç organizasyonu, alt bileşenleri ve nörobiyolojik karşılıkları ayrıntılı biçimde tanımlanmıştır. Bu bölümde ise, aynı yapıları yeniden yapısal açıdan değil, *işlevsel açıdan* bakıyoruz. Amaç,  $M_t$  ve  $W_t$ 'nin içeriğinin değer fonksiyonuna nasıl yansıdığını, yani bellek ve öncelik bileşenlerinin çöküş mekanizması içinde hangi matematiksel rolleri üstlendiğini formel olarak tanımlamaktır. Dolayısıyla 3.2 yapısal temeli, Bu bölümde ise bu temel üzerinde çalışan işlevsel değerlendirme katmanını oluşturur.

Jeneratif operatör  $G$ , her zaman adımında bilinç için olası aday temsiller kümesini

$$X_t = \{x_t^{(1)}, \dots, x_t^{(K)}\}$$

üretir (bkz. Bölüm 3.4.1). Ancak bu adaylar, eşit derecede işlevsel veya olası değildir. FBÇT'de çöküş operatörü  $C_t$ , her adayı çok bileşenli bir *değer fonksiyonu* ile değerlendirir ve bu değerlere bağlı bir olasılık dağılımı üzerinden seçim yapar. Bu bölümde  $f$  fonksiyonunun yapısını ve bileşenlerini tanımlıyoruz.

#### 3.5.1 Temel tanım ve ayrıştırma

Her  $x \in X$  adayı için, zaman  $t$ 'deki değer fonksiyonu

$$f_t : X \times S \times M \times W \rightarrow \mathbb{R}$$

şu şekilde tanımlanır:

$$f_t(x) \equiv f(x; S_t, M_t, W_t).$$

FBÇT, bu fonksiyonun üç temel bileşene ayrıştırılabileceğini varsayar:

$$f(x; S_t, M_t, W_t) = \alpha_{S,t} f_S(x; S_t) + \alpha_{M,t} f_M(x; M_t) + \alpha_{W,t} f_W(x; W_t),$$

burada  $\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t} \geq 0$ , farklı kaynakların göreceli önemini belirleyen ağırlıklardır (Ayrıntılar için bkz. Bölüm 3.6).

Sezgisel olarak:

- $f_S$  duyusal uygunluğu,
- $f_M$  bellek uyumunu,
- $f_W$  ise değer/motivasyon uyumunu temsil eder.

### 3.5.2 Duyusal bileşen: $f_S$

Duyusal bileşen, aday  $x$ 'in mevcut duyusal girdiyi ne kadar iyi açıkladığını ölçer. Bölüm 3.4.1'de tanımlanan duyusal enerji terimi

$$E_S(x; S_t) = \frac{1}{2\sigma_S^2} \|x - \phi_S(S_t)\|^2$$

üzerinden, doğal bir değer bileşeni elde ederiz:

$$f_S(x; S_t) = -E_S(x; S_t) = -\frac{1}{2\sigma_S^2} \|x - \phi_S(S_t)\|^2.$$

Böylece duyusal temsile yakın adaylar daha yüksek  $f_S$  alır; bu, yüksek duyusal uyumun çöküş sırasındaki etkisini açıklar.

### 3.5.3 Bellek bileşeni: $f_M$

Bellek bileşeni, aday  $x$ 'in  $M_t$ 'de kodlu geçmiş deneyimlerle ne kadar uyumlu olduğunu ölçer. Jeneratif düzeyde kullanılan

$$E_M(x; M_t) = \mathbb{E}_{m \sim M_t} \left[ \frac{1}{2\sigma_M^2} \|x - \phi_M(m)\|^2 \right]$$

enerji terimine paralel olarak:

$$f_M(x; M_t) = -E_M(x; M_t) = -\mathbb{E}_{m \sim M_t} \left[ \frac{1}{2\sigma_M^2} \|x - \phi_M(m)\|^2 \right].$$

Bu, adayın bellekteki benzer durumlarla ne derece “tanıdık” olduğunu nicelleştirir. Yüksek  $f_M$  alan adaylar, geçmiş deneyimlerle daha tutarlı yorumlara karşılık gelir.

### 3.5.4 Değer bileşeni: $f_W$

Değer bileşeni, adayın mevcut öncelik yapısıyla (örn. hedefler, tehdit, merak) ne kadar uyumlu olduğunu ölçer. Bölüm 3.4.1'de tanımlanan

$$V(x; W_t) = \lambda_W W_t^\top \psi(x)$$

potansiyeline paralel olarak,  $f_W$ 'yu basitçe

$$f_W(x; W_t) = W_t^\top \psi(x)$$

şeklinde tanımlarız (ölçek faktörü  $\lambda_W$  ağırlık  $\alpha_{W,t}$  içine katılabilir). Böylece, değer vektörü ile aynı yönde projekte olan adaylar daha yüksek  $f_W$  alır; bu, davranışsal olarak avantajlı yorumların çöküşte öne çıkmasını sağlar.

### 3.5.5 Ölçeklendirme ve normalizasyon

Farklı bileşenlerin doğal ölçekleri birbirinden çok farklı olabilir. Bu durum, tek bir bileşenin (örneğin  $f_W$ ) diğerlerini domine etmesine ve  $\alpha$  ağırlıklarının işlevsiz hale gelmesine yol açabilir. Bu nedenle her bileşen, zaman içinde izlenen istatistiklerle normalize edilir:

$$\begin{aligned} \tilde{f}_S(x; S_t) &= \frac{f_S(x; S_t) - \mu_{S,t}}{\sigma_{S,t} + \varepsilon}, & \tilde{f}_M(x; M_t) &= \frac{f_M(x; M_t) - \mu_{M,t}}{\sigma_{M,t} + \varepsilon}, \\ \tilde{f}_W(x; W_t) &= \frac{f_W(x; W_t) - \mu_{W,t}}{\sigma_{W,t} + \varepsilon}, \end{aligned}$$

burada  $\mu_{(\cdot),t}$  ve  $\sigma_{(\cdot),t}$  zaman içinde tutulan kayan ortalama ve standart sapmayı,  $\varepsilon > 0$  ise sayısal kararlılığı sağlayan küçük bir sabiti temsil eder.

Bu normalize edilmiş biçimleri kullanarak:

$$f(x; S_t, M_t, W_t) = \alpha_{S,t} \tilde{f}_S(x; S_t) + \alpha_{M,t} \tilde{f}_M(x; M_t) + \alpha_{W,t} \tilde{f}_W(x; W_t).$$

### 3.5.6 Somut örnek

Bölüm 3.4.1'deki belirsiz bahçe senaryosunu yeniden ele alalım.  $G$  operatörü aşağıdaki adayları üretmiş olsun:

$$\begin{aligned} x^{(1)} &: \text{"hırsız"}, \\ x^{(2)} &: \text{"komşu"}, \\ x^{(3)} &: \text{"kedi"}, \\ x^{(4)} &: \text{"belirsiz gölge"}. \end{aligned}$$

**Duyusal bileşen.** Düşük çözünürlüklü gölgesel bir hareket için, hem "hırsız", hem "komşu", hem de "kedi" temsilleri, benzer bir silüetle uyumlu olabilir. Bu nedenle  $f_S$  değerleri birbirine görece yakın olabilir; "belirsiz gölge" adayı ise daha nötr bir yorum olduğu için duyusal açıdan da makul kalır:

$$f_S(x^{(1)}) \approx f_S(x^{(2)}) \approx f_S(x^{(3)}) \gtrsim f_S(x^{(4)}).$$

**Bellek bileşeni.** Kişi, geçmişte sıkça kediye bahçede görmüş, komşunun da geç saatlerde geldiğine aşınadır; oysa hırsızlık olayı nadirdir. Bu durumda:

$$f_M(x^{(3)}) > f_M(x^{(2)}) > f_M(x^{(1)}),$$

yani bellek açısından "kedi" en tanıdık yorumdur.

**Değer bileşeni.** Buna karşın,  $W_t$ 'deki tehdit bileşeni yüksektir; bu nedenle  $\psi(x^{(1)})$ 'in tehdit boyutu yüksektir ve

$$f_W(x^{(1)}) \gg f_W(x^{(3)}), f_W(x^{(2)}), f_W(x^{(4)}).$$

**Toplam değer.** Eğer bağlamda  $\alpha_{W,t}$  görel olarak yüksekse (örneğin kişi zaten gergin, yalnız ve risk odaklıysa), toplam değer:

$$f(x^{(1)}) > f(x^{(2)}), f(x^{(3)}), f(x^{(4)})$$

şeklinde sıralanabilir ve çöküş operatörü  $C_t$ , "hırsız" yorumunu daha yüksek olasılıkla seçer. Daha güvenli veya rahat bir bağlamda (düşük  $\alpha_{W,t}$ , yüksek  $\alpha_{M,t}$ ) ise "kedi" veya "komşu" yorumları öne çıkar. Böylece aynı duyusal girdiden çok farklı bilinçli deneyimlerin ("tehdit", "masumiyet") nasıl türediği fonksiyonel düzeyde açıklanmış olur.

### 3.5.7 Çöküş operatörü ile entegrasyon

Çöküş operatörü, bu değer fonksiyonunu kullanarak adaylar üzerinde Boltzmann-softmax biçiminde bir olasılık dağılımı tanımlar:

$$\pi_t(x) = \frac{\exp(\beta_t f(x; S_t, M_t, W_t))}{\sum_{j=1}^K \exp(\beta_t f(x_t^{(j)}; S_t, M_t, W_t))}.$$

Bilinçli içerik daha sonra bu dağılımdan örnekleme ile seçilir:

$$C_t = \text{sample}[\pi_t].$$

Böylece, 3.2'de tanımlanan bellek ve öncelik sistemleri, 3.5'te tanımlanan değer fonksiyonu aracılığıyla çöküş mekanizmasına doğrudan bağlanmış olur.

### 3.5.8 Çöküş Operatörünün Zorunluluğu

Fonksiyonel Bilinç Çöküşü Teorisi'nde (FBÇT) çöküş operatörü

$$C : (S_t, M_t, W_t) \mapsto C_t,$$

basit bir seçim mekanizması değildir; teorinin matematiksel, bilişsel ve nörobiyolojik yapısından zorunlu olarak türetilir. Bu bölümde, çöküş fonksiyonunun neden Boltzmann-softmax biçiminde olması gerektiğini ve alternatif seçim operatörlerinin neden FBÇT ile uyumsuz sonuçlar verdiğini gösteriyoruz.

**Tekillik ilkesi: Bilinç tek bir içeriğe projekte olmak zorundadır.** Fenomenolojik olarak bilinç, aynı anda tek bir içeriği küresel erişime açar. Bu “global availability” ilkesi [18, 19], matematiksel olarak bir projeksiyon operatörünü gerektirir:

$$C : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{X}.$$

Dolayısıyla jeneratif uzayda  $K$  tane olası aday olsa bile, bilinç seviyesi tek bir temsile indirgenmek zorundadır.

**Değerleme fonksiyonunun zorunluluğu.** Her bilinç adımında aday durumları karşılaştırmak için sürekli, türevlenebilir, monoton bir değer fonksiyonuna ihtiyaç vardır:

$$f : X \times S \times M \times W \rightarrow \mathbb{R}.$$

Bu fonksiyon, duyuşal uygunluk ( $f_S$ ), bellek uyumluluğu ( $f_M$ ) ve motivasyonel değer ( $f_W$ ) bileşenlerinin ortak bir ölçeğe aktarılmasını sağlar. Öğrenme dinamiklerinin (Eş. 74–77) türevlenebilirlik gereksinimi,  $f$ 'nin en az  $C^1$  olmasını zorunlu kılar.

**Deterministik argmax'ın yetersizliği.**  $C_t = \arg \max f(x)$  biçimindeki deterministik bir seçim:

(i) aynı değere sahip adaylar arasında simetriyi kıramaz, (ii) duyuşal gürültü altında gözlenen mikro-zamansal fluktuasyonları açıklayamaz, (iii) öğrenme sinyallerini boğarak geri beslemeyi devre dışı bırakır.

Dolayısıyla bilinç anı deterministik olamaz; seçim *değer-duyarlı fakat stokastik* olmalıdır.

**Yönlendirilmiş rastlantısallık: Ekspansiyel familyanın zorunluluğu.** Seçim mekanizmasının şu üç özelliği aynı anda sağlanması gerekir:

1. **Değer duyarlılığı:**  $f(x)$  arttıkça seçilme olasılığı artmalıdır.
2. **Öğrenilebilirlik:** Türevlenebilir olmalı,  $\nabla f$  üzerinden güncellenebilir olmalıdır.
3. **Normalizasyon:** Olasılıkların toplamı 1 olmalıdır.

Bu üç koşulu aynı anda sağlayan *tek* dağılım ailesi, maksimum entropi prensibine göre ekspansiyel formdur:

$$\pi_t(x) \propto \exp(\beta_t f(x, S_t, M_t, W_t)).$$

**Sıcaklık parametresi  $\beta_t$ 'nin zorunluluğu.**  $\beta_t$ , belirsizlik düzeyini ve seçim keskinliğini kontrol eden bilişsel bir “gain” parametresidir:

$$\beta_t \rightarrow 0 : \text{keşif (yüksek entropi)}, \quad \beta_t \rightarrow \infty : \text{keskin, neredeyse deterministik seçim}.$$

Bu yapı, noradrenerjik ve dopaminerjik sistemlerin davranışsal kararlardaki rolüyle [20, 21] nörobiyolojik olarak doğrulanmaktadır.

**Alternatif çöküş mekanizmalarının elenmesi.** Aşağıdaki operasyonel formlar FBÇT ile uyumsuzdur:

- **Lineer seçim:**  $\pi \propto f$  ölçek değişmezliği ve normalizasyon sorunları doğurur.
- **Winner-take-all (WTA):** Türevlenebilir değildir;  $C$  öğrenmeyi durdurur ve  $\Delta M_t - \Delta W_t$  güncellemeleri çalışamaz hale gelir.
- **Tam rastgele seçim:** Değerleme tamamen ihmal edilir; bilinç deneyimi ile davranış arasında gözlenen korelasyon açıklanamaz.
- **Deterministik argmax:** Mikro-varyasyonlar kaybolur, soft attention mekanizmaları modelden çıkmış olur.

**FBÇT içindeki çöküşün işlevsel rolü.** Çöküş yalnızca o anki bilinç içeriğini üretmez; aynı zamanda gelecekteki güncellemelerin kaynağıdır:

$$(M_{t+1}, W_{t+1}) = F(M_t, W_t, S_t, C_t).$$

Bu geri-besleme döngüsü sayesinde:

- her seçim bellek izleri bırakır (path dependence),
- öncelikler yeniden şekillenir,
- uzun ölçekli kimlik ve kişilik sürekliliği ortaya çıkar,
- $\alpha$  ağırlıkları bağlamsal olarak yeniden düzenlenir.

**Sonuç olarak, Boltzmann-softmax formu matematiksel bir zorunluluktur.** FBÇT'nin yapısal ilkeleri, (i) teklik, (ii) değer-duyarlı stokastik seçim, (iii) enerji tabanlı tutarlılık, (iv) öğrenilebilirlik, (v) nörobiyolojik uyumluluk

koşullarını aynı anda sağlar. Bu koşulları aynı çerçevede karşılayan *tek* form:

$$C_t = \text{sample} \left[ \frac{\exp(\beta_t f(x))}{\sum_j \exp(\beta_t f(x_j))} \right]$$

şeklindeki Boltzmann-softmax operatörüdür.

Dolayısıyla çöküş fonksiyonu, FBÇT'de keyfi bir seçim değil, teoremin yapısından zorunlu olarak türetilen tek matematiksel formdur.

### 3.6 Bağlamsal Ağırlıkların Öğrenilmesi ve Dinamik Güncellenmesi

**Motivasyon.** Önceki bölümde değer fonksiyonu  $f$ 'nin üç bilgi kaynağını (duyusal uygunluk, bellek uyumu, motivasyonel değer) bir araya getirdiği gösterildi. Ancak bu kaynakların göreceli önemi sabit kabul edilemez; gerçek bilişsel sistemlerde bu öncelikler bağlama göre sürekli değişir. Örneğin tehdit altında değer bileşeni ağır basarken, tanıdık ortamlarda bellek bileşeni baskın hale gelir. Bu nedenle FBÇT, ağırlıkların sabit katsayılar değil, bağlama göre öğrenilen ve değişen dinamik değişkenler olduğunu varsayar.

Önceki bölümde tanımlanan değer fonksiyonu

$$f(x; S_t, M_t, W_t) = \alpha_{S,t} \tilde{f}_S(x; S_t) + \alpha_{M,t} \tilde{f}_M(x; M_t) + \alpha_{W,t} \tilde{f}_W(x; W_t)$$

bileşen skorlarının  $\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t}$  ağırlıkları ile birleştirilmesine dayanır. Bu alt bölümde, bu ağırlıkların yalnızca sabit katsayılar değil, bağlama duyarlı dinamik değişkenler olduğunu formel hale getiriyoruz.



### 3.6.1 Temel kısıtlar ve yorum

Ağırlıklar, her zaman adımında şu kısıtları sağlar:

$$\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t} \geq 0, \quad \alpha_{S,t} + \alpha_{M,t} + \alpha_{W,t} = 1.$$

Dolayısıyla  $(\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t})$  üçlüsü, üç boyutlu bir olasılık simplex'i üzerinde yaşayan bir bağlam vektörüdür. Sezgisel olarak:

- $\alpha_{S,t}$  yüksek olduğunda sistem *duyusal veriye daha çok güvenir*; hızlı, veri-odaklı yorumlar baskındır.
- $\alpha_{M,t}$  yüksek olduğunda sistem *geçmiş deneyimlere yaslanır*; tanıdıklık ve alışkanlık ağır basar.
- $\alpha_{W,t}$  yüksek olduğunda sistem *değer ve hedef yapısı tarafından yönlendirilir*; motivasyonel önyargılar belirginleşir.

### 3.6.2 Bağlam vektörü ve parametrizasyon

Ağırlıkları doğrudan modellemek yerine, önce bağlamı özetleyen bir özellik vektörü tanımlarız:

$$c_t = (u_{S,t}, u_{M,t}, u_{W,t}, r_t, \xi_t) \in \mathbb{R}^d.$$

Burada örnek olarak:

- $u_{S,t}$ : duyusal belirsizlik ölçüsü (örneğin, tahmin hatası varyansı),
- $u_{M,t}$ : bellek güvenilirlik ölçüsü (örneğin, bellek temsillerinin tutarlılığı),
- $u_{W,t}$ : değer gradyanının büyüklüğü (örn. beklenen ödülde değişim),
- $r_t$ : son dönemde alınan toplam ödül/başarım sinyali,
- $\xi_t$ : fizyolojik/durum değişkenleri (uyku hali, stres vb.).

Ağırlıklar, bu bağlam vektörü üzerinden parametrik bir fonksiyonla tanımlanır:

$$\alpha_t = (\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t}) = \text{softmax}(g_\Theta(c_t)),$$

burada  $g_\Theta : \mathbb{R}^d \rightarrow \mathbb{R}^3$  parametrik bir haritalamadır (örneğin küçük bir sinir ağı) ve softmax operatörü simplex kısıtlarını otomatik olarak sağlar:

$$\alpha_{i,t} = \frac{\exp(z_{i,t})}{\sum_j \exp(z_{j,t})}, \quad \mathbf{z}_t = g_\Theta(c_t).$$

### 3.6.3 Öğrenme kuralı

$\Theta$  parametreleri, sistemin uzun vadeli başarısını yansıtan bir hedef fonksiyon üzerinden öğrenilir. Örneğin, bilinçli durumların neden olduğu davranışların başarısını ölçen bir ödül sinyali  $R_t$  verildiğinde:

$$\Theta_{t+1} = \Theta_t + \eta_\alpha \nabla_\Theta \mathbb{E}[R_t | \Theta_t],$$

veya pratikte, uygun bir kayıp fonksiyonu  $\mathcal{L}_\alpha$  üzerinden yaklaşık gradyan inişi yapılır:

$$\Theta_{t+1} = \Theta_t - \eta_\alpha \nabla_\Theta \mathcal{L}_\alpha(\Theta_t).$$

Bu, sistemin zaman içinde hangi bağlamlarda duyusal veriye, hangi bağlamlarda ise belleğe veya değerlere daha çok ağırlık vermesi gerektiğini deneyim üzerinden “öğrenmesini” sağlar.

### 3.6.4 Örnek: tehdit bağlamında ağırlık kayması

Belirsiz bahçe senaryosuna (Bölüm 3.4.1 ve 3.5) geri dönelim. İki uç bağlam düşünelim:

- **Güvenli bağlam:** Kişi genelde güvende hissetmekte, bölgede son dönemde tehdit yaşamamış,  $r_t$  olumlu ve  $u_{S,t}$  orta düzeydedir. Öğrenilmiş parametreler sonucunda

$$\alpha_t^{(\text{güvenli})} \approx (0.3, 0.5, 0.2),$$

yani bellek ağırlığı baskındır; “kedi” ve “komşu” gibi tanıdık yorumlar öne çıkar.

- **Tehdit bağlamı:** Son dönemde hırsızlık yaşanmış, genel güvenlik algısı düşmüş, stres ve uyarılma düzeyi artmıştır;  $u_{S,t}$  (belirsizlik) ve  $u_{W,t}$  (tehditle ilişkili değer gradyanı) yüksektir. Bu durumda

$$\alpha_t^{(\text{tehdit})} \approx (0.2, 0.2, 0.6),$$

yani değer bileşeni baskın hale gelir; “hırsız” yorumuna karşılık gelen adaylar çöküşte daha yüksek olasılık kazanır.

Bu örnek, aynı duyuşsal input ve benzer bellek içeriği altında bile  $\alpha_t$ ’nin bağlama göre değişmesinin bilinçli deneyimi nasıl radikal biçimde dönüştürebildiğini gösterir.

### 3.6.5 Nöromodülatör yorum

Ağırlık dinamikleri, nörobiyolojik düzeyde nöromodülatör sistemlerle doğal biçimde ilişkilendirilebilir:

- Noradrenerjik ton (loкус seruleus) artışı, genellikle artmış uyarılma ve tehdit değerlendirme-siyle ilişkilidir; bu, modelde  $\alpha_{W,t}$ ’nin yükselmesine karşılık gelebilir.
- Serotonerjik ton, uzun vadeli istikrar ve alışkanlık süreçleriyle ilişkilidir; bu,  $\alpha_{M,t}$  bileşeninin gö-reli ağırlığını artırabilir.
- Kolinerjik modülasyon, duyuşsal dikkat ve giriş işlenmesini güçlendirir; bu,  $\alpha_{S,t}$ ’nin yüksek olduğu durumlarla örtüşür.

Bu yorumlar, FBÇT’nin yalnızca fonksiyonel bir model değil, aynı zamanda nörobiyolojik hipotezler üreten bir çerçeve olarak kullanılabileceğini göstermektedir.

### 3.6.6 Sınır durumlar ve özel rejimler

Bazı sınır durumlar, bilincin farklı fenomenolojik rejimlerine karşılık gelir:

- $\alpha_{S,t} \approx 1$ : *Anlık, duyuşsal ağırlıklı bilinç.* Duyusal akış baskındır, içsel yorumlar ve hedefler geri plandadır.
- $\alpha_{M,t} \approx 1$ : *Geçmişe yaslı bilinç.* Ruminasyon, hatırlama ve çağrışıma dayalı içsel akışlar baskındır.
- $\alpha_{W,t} \approx 1$ : *Hedef/tehdit odaklı bilinç.* Motivasyonel ve duyuşsal içerikler baskındır; aynı duyuşsal input, sık sık tehdit veya fırsat olarak yorumlanır.

Bu rejimler, bilinçli deneyimin farklı klinik ve psikolojik durumlarla (anksiyete, depresyon, travma sonrası tepkiler gibi) nasıl ilişkilenebileceğine dair daha ileri tartışmalar için bir başlangıç noktası sağlar (bkz. Bölüm 6.6).

## 3.7 Çöküş Mekanizması

FBÇT’nin merkezinde, aday dağılımından bilinçli durum seçimini gerçekleştiren bir çöküş mekanizması bulunur. Bu mekanizma, aday üretim operatörü  $\mathcal{G}$ , değerlendirme operatörü  $\mathcal{E}$  ve seçim/örnekleme operatörü  $\mathcal{K}$ ’nin bileşimi olarak tanımlanır.

### 3.7.1 Öz Model Sabit Nokta Koşulu

FBÇT’de öz-bilinç durumu, içsel temsil uzayının dinamik bir sabit nokta etrafında kararlılık kazanmasıyla tanımlanır. Bu yapı, karar vericinin kendisi hakkındaki temsiliğin stabil bir biçimde sürdürülebilmesini sağlar.

**Tanım 3.1** (Self-Model Fixed Point). Bir öz-temsil  $x_{\text{self}} \in X_{\text{self}}$ , aşağıdaki koşulu sağlıyorsa *sabit nokta* olarak tanımlanır:

$$x_{\text{self}}^* = G_{\text{self}}(S_t, M_t, W_t, x_{\text{self}}^*), \quad (60)$$

burada  $G_{\text{self}}$ ,  $G$  generatif fonksiyonunun öz-model bileşenini temsil eder.

Bu sabit nokta, öz-modelin dış girdilerle birlikte kendi önceki hâlimden de beslendiğini gösterir. Stabilite koşulu:

$$\|G_{\text{self}}(S_t, M_t, W_t, x) - G_{\text{self}}(S_t, M_t, W_t, x^*)\| \leq \lambda \|x - x^*\|, \quad \lambda < 1, \quad (61)$$

$G_{\text{self}}$  fonksiyonunun kontraktif olduğunu ve sabit noktanın iyi tanımlı olduğunu garanti eder. Bu yapı, öz-düşünümün ve öznel sürekliliğin matematiksel temsiliğini oluşturur.

**Fenomenolojik yorum.** Bu sabit nokta koşulu, "ben kimim?" sorusunun dinamik cevabını temsil eder. Dissociation ve depersonalization bozukluklarında (6.6, Tahmin 16), bu sabit nokta sistematik olarak ihlal edilir:  $\|x_{\text{self}} - x_{\text{self}}^*\| > \varepsilon$ .

### 3.7.2 Tam çöküş operatörü

Tam çöküş operatörü

$$\mathcal{C}_t : \mathcal{S} \times \mathcal{M} \times \mathcal{W} \rightarrow X \quad (62)$$

şu şekilde verilir:

$$\mathcal{C}_t = \mathcal{K}_{\beta_t} \circ \mathcal{E} \circ \mathcal{G}, \quad (63)$$

ve bilinç durumu

$$C_t = \mathcal{C}_t(S_t, M_t, W_t) \in X \quad (64)$$

şeklinde tanımlanır.

Burada

- $\mathcal{G}(S_t, M_t, W_t) = \mu_t$ : aday dağılımını üreten operatör,
- $\mathcal{E}$ :  $\mu_t$  üzerindeki adaylara  $f$  veya  $\mathbb{V}_t$  üzerinden skor atayan değerlendirme operatörü,
- $\mathcal{K}_{\beta_t}$ : inverse sıcaklık parametresi  $\beta_t$  ile parametrize edilen olasılıksal seçim (örnekleme) operatörü.

Bu kompozisyon, çöküşün ne tamamen rastlantısal (ne saf gürültü), ne de tamamen deterministik olduğunu; fakat değer fonksiyonuna göre şekillenmiş, kontrol edilebilir bir stokastik seçim olduğunu garanti eder.

**Bilinç Seviyesi Ölçütünün Matematiksel Özellikleri** Tanım gereği:

$$L_t = D_{\text{KL}}(\pi_t \| P_t), \quad (65)$$

bilinç durumunun normatif beklentiden sapma derecesini ölçer. Bu metrik aşağıdaki temel özelliklere sahiptir.

#### 1. Negatifsizlik

$$L_t \geq 0.$$

Bu, bilinç düzeyinin hiçbir zaman negatif olamayacağını gösterir.

## 2. Sıfır eşdeğerliği

$$L_t = 0 \iff \pi_t = P_t.$$

Bu koşul, dikkat dağınıklığı veya otomatik davranışlarda  $L_t$ 'nin minimum değere yakın olmasını açıklar.

## 3. Monotonluk Seçim dağılımı keskinleştikçe entropi azalır ve

$$L_t \propto \frac{1}{H(\pi_t)}.$$

## 4. Sonlu durum uzayında sınırlılık Eğer aday sayısı $n$ ise:

$$0 \leq L_t \leq \log n.$$

## 5. Nörobiyolojik yorumlanabilirlik Yüksek $L_t$ , frontoparietal entegrasyon *artışı* ve sensörimotor segregasyon *azalması* ile ilişkilidir.

### 3.7.3 Politika dağılımı

Aday dağılımı  $\mu_t$  ve skor fonksiyonu  $f$  kullanılarak bir politika (seçim) dağılımı tanımlanır:

$$\pi_t(dx) = \frac{\exp(\beta_t f(x, S_t, M_t, W_t))}{Z_t} \mu_t(dx), \quad (66)$$

burada

$$Z_t = \int_X \exp(\beta_t f(x, S_t, M_t, W_t)) \mu_t(dx) \quad (67)$$

normalizasyon sabitidir ve  $\beta_t > 0$  *inverse sıcaklık* parametresidir. Büyük  $\beta_t$  değerleri daha deterministik, küçük  $\beta_t$  değerleri daha keşifçi (stochastic) politika davranışına karşılık gelir.

Ayrık aday kümesi için  $x_t^{(i)} \sim \mu_t, i = 1, \dots, N$ :

$$\pi_t^{(i)} = \frac{\exp(\beta_t f^{(i)})}{\sum_{j=1}^N \exp(\beta_t f^{(j)})}, \quad f^{(i)} = f(x_t^{(i)}, S_t, M_t, W_t). \quad (68)$$

### 3.7.4 Bilinç durumunun seçimi

Bilinç durumu, politika dağılımından örneklenen bir çöküş olarak tanımlanır:

$$C_t \sim \pi_t(dx), \quad C_t \in X. \quad (69)$$

Bu  $C_t$ ,

- sistemin o anki içsel bilişsel durumunu,
- dışa vurulan davranışı/kararı,
- ve fenomenolojik deneyimin taşıyıcısını

birlikte temsil eder. Teorinin iddiası, “bilinç anı”nın tam olarak bu çöküş işlemiyle özdeşleştirilebileceğidir: bilinç, belirli bir  $C_t$  örneğinin gerçekleşmiş olmasıdır.

## 3.8 Geri Bildirim ve Güncelleme Dinamiği

FBÇT, tekil çöküş anlarını izole olaylar olarak değil, zaman içinde  $M_t$  ve  $W_t$  üzerinde geribildirim oluşturan bir süreç olarak ele alır. Böylece bilinç yalnızca anlık bir seçimi değil, kendi geçmişini şekillendiren bir *öğrenme dinamiği*ni de içerir.

**Genel güncelleme denklemi.**

$$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t), \quad (70)$$

burada  $\mathcal{F}$ , öğrenme ve adaptasyon dinamiklerini topluca temsil eden bir operatördür.

Tipik bir ayrıştırma:

$$M_{t+1} = M_t + \Delta M_t(S_t, C_t, M_t), \quad (71)$$

$$W_{t+1} = W_t + \Delta W_t(S_t, M_t, C_t, R_t), \quad (72)$$

$R_t$  uygun bir ödül/ceza veya başarımlı sinyalidir.

**Bellek güncellemesi.**

$$\Delta M_t = \eta_M \cdot h(C_t, S_t, M_t), \quad (73)$$

burada  $\eta_M > 0$  öğrenme hızıdır. İki tipik örnek:

**Hebbian-benzeri:**

$$\Delta M_t = \eta_M [C_t \otimes \phi(S_t) - \gamma M_t], \quad (74)$$

tekrarlanan deneyimlerin güçlendirilmesi ( $C_t$  ile  $S_t$ 'nin birlikte etkinliği) ve  $\gamma > 0$  ile modellenen unutkanlık dinamiğini içerir.

**Tahmin hatası tabanlı:**

$$\Delta M_t = \eta_M [C_t - \hat{C}_t(M_t, S_t)], \quad (75)$$

prediktif kodlama perspektifinde, beklenen  $\hat{C}_t$  ile gerçekleşen  $C_t$  arasındaki farkı minimize etmeye çalışan bir güncellemeyi ifade eder.

**Ağırlık vektörü güncellemesi.** Öncelik vektörü, ödül/ceza sinyali  $R_t$  üzerinden güncellenir:

$$\Delta W_t = \alpha_W \cdot g(R_t, W_t, C_t), \quad (76)$$

burada  $\alpha_W > 0$  adaptasyon hızıdır.

**Gradyan bazlı:** Eğer sistem belirli bir hedef fonksiyonunu optimize ediyorsa:

$$\Delta W_t = \alpha_W \nabla_W \mathbb{E}_{\pi_t}[R(C_t, S_t)], \quad (77)$$

ve simplex kısıtını korumak için:

$$W_{t+1} = \Pi_{\Delta^{k-1}}(W_t + \Delta W_t), \quad (78)$$

burada  $\Pi_{\Delta^{k-1}}$  simplex üzerine izdüşüm operatörüdür (örneğin softmax normalizasyonu).

**Homeostatik düzenleme:** Sistemin içsel dengesini koruması gereken durumlarda:

$$\Delta W_t = \alpha_W [W^{\text{target}} - W_t], \quad (79)$$

burada  $W^{\text{target}}$  biyolojik veya görev gereksinimleri tarafından belirlenen hedef ağırlık vektörüdür.

**Çok ölçekli dinamikler.** Güncellemeler farklı zaman ölçeklerinde gerçekleşir:

- **Hızlı** ( $\sim \text{ms-s}$ ): Dikkat, nöromodülasyon ( $\eta_M \sim 10^{-1}$ )
- **Orta** ( $\sim \text{dakika-saat}$ ): Strateji, alışkanlık ( $\eta_M \sim 10^{-3}$ )
- **Uzun** ( $\sim \text{gün-yıl}$ ): Kişilik, değerler ( $\eta_M \sim 10^{-5}$ )

FBQT bu ölçekleri birleştirir:

$$\Delta M_t = \sum_{\tau} \eta_M^{(\tau)} h^{(\tau)}(C_t, S_t, M_t), \quad (80)$$

burada  $\tau \in \{\text{fast, medium, slow}\}$  farklı zaman ölçeklerini temsil eder.

Bu yapı, bilinçteki tarihsel bağımlılığı ve kademeli değişimleri (alışkanlık, travma etkisi, uzun vadeli adaptasyon) doğal bir şekilde modelleyebilir.

### 3.9 Benlik Modeli ve Sabit Nokta Koşulu

Teorinin ileri düzey bileşenlerinden biri, benlik temsilinin  $X_t$  içindeki rolüdür. Burada amaç, benlik-modeli ile bilinç durumu arasındaki ilişkiyi fonksiyonel bir sabit nokta koşulu ile formelleştirmektir.

**Benlik gözlem operatörü.** Önce, sistemin kendi durumu hakkında erişebildiği bilgiyi temsil eden bir gözlem operatörü tanımlanır:

$$H_t = H(S_t, M_t, W_t) \in \mathcal{H}, \quad (81)$$

burada  $\mathcal{H}$ , *benlik hakkında erişilebilir bilgi* uzayıdır (örneğin beden pozisyonu, duygusal durum, bilişsel hedefler, geçmiş eylemler vb.).

**Benlik üretici fonksiyon.** Benlik modeli, bu gözlemlerden türetilen bir haritalama ile temsil edilir:

$$F_{\text{self}} : \mathcal{H} \rightarrow \mathbb{R}^{d_s}, \quad \hat{x}_t^{\text{self}} = F_{\text{self}}(H_t). \quad (82)$$

**Sabit nokta koşulu.** FBÇT, benlik bileşeninin yaklaşık bir sabit nokta oluşturduğunu varsayar:

$$\|x_t^{\text{self}} - F_{\text{self}}(H_t)\| \leq \varepsilon, \quad (83)$$

burada  $\varepsilon \geq 0$  küçük bir toleranstır. Bu koşul sağlandığında, sistemin o anda sahip olduğu benlik temsili ile kendi hakkında ulaşabildiği bilgi arasında yüksek düzeyde tutarlılık olduğu söylenebilir.

**Değer fonksiyonuna entegrasyon.** Benlik tutarlılığının bilinçli seçim üzerinde rol oynaması için skor fonksiyonuna bir ceza terimi eklenebilir:

$$f_{\text{tot}}(x, S_t, M_t, W_t) = \langle W_t, U(x, S_t, M_t) \rangle - \lambda \|x^{\text{self}} - F_{\text{self}}(H_t)\|^2, \quad (84)$$

$\lambda \geq 0$  benlik-tutarlılık ağırlığıdır. Böylece, benlik modelini bozacak aday durumlar, fonksiyonel olarak dezavantajlı hâle gelir.

Bu yaklaşım, dissociation ve depersonalization bozukluklarının formal tanımına olanak tanır: bu durumlarda (83) koşulu sistematik olarak ihlal edilir ve klinik fenomenoloji tam da bu sabit nokta kaybının bir sonucu olarak yorumlanabilir.

### 3.10 Bilinç Düzeyi Metriği

FBÇT, yalnızca hangi durumun bilinçli olduğuna değil, aynı zamanda o anki *bilinç düzeyinin* nicel olarak nasıl ölçülebileceğine de bir öneri sunar.

**Referans dağılım.** Önce, aynı zamanda  $(S_t, M_t, W_t)$  için referans kabul edilen bir dağılım  $P_t$  tanımlanır. Bu dağılım,

- sistemin tipik/varsayılan politika davranışını,
- veya uzun dönem ortalama seçim dağılımını

temsil edebilir:

$$P_t(dx) = P(X_t \in dx \mid S_t, M_t, W_t). \quad (85)$$

**Bilinç düzeyi ölçütü.** Bilinç düzeyi  $L_t$ , politika dağılımı  $\pi_t$  ile referans dağılım  $P_t$  arasındaki Kullback-Leibler ıraksaması olarak tanımlanır:

$$L_t = D_{\text{KL}}(\pi_t \| P_t) = \int_X \log \frac{\pi_t(x)}{P_t(x)} \pi_t(dx). \quad (86)$$

Alternatif olarak, entropi farkı formülasyonu:

$$L_t = H(P_t) - H(\pi_t), \quad (87)$$

burada  $H(\cdot)$  Shannon entropisidir.

**Yorum.**

- $L_t$  yüksek ise,  $\pi_t$  referansa göre keskinleşmiş ve belli bir alt-uzaya yoğunlaşmış demektir; bu, güçlü odaklanma ve seçim seçiciliği ile ilişkilendirilebilir.
- $L_t$  düşük ise,  $\pi_t$  referans dağılıma yakın, dağınık ve seçici olmayan bir profil sergiler; bu, düşük uyarılmışlık ve zayıf bilinç düzeyiyle uyumludur.

Bu metrik, farklı bilinç hallerinin (uyanıklık, uyku, anestezi, yoğun dikkat, meditasyon vb.) nicel olarak karşılaştırılmasına olanak tanır ve deneysel validasyon için somut bir hedef sunar (bkz. 6.2, Tahmin 15).

### 3.11 Qualia'nın Matematiksel Yapısı

FBÇT, fenomenolojik içeriği, iç temsil uzayı ile duyusal durum uzayının üzerinde tanımlanan bir eşdeğerlik sınıfı yapısı olarak formalize eder.

**Eşdeğerlik ilişkisi.**

$$(x_1, S_1) \sim (x_2, S_2) \quad (88)$$

eğer ve yalnızca eğer sistem, tüm erişilebilir fonksiyonel ayrımlar bakımından bu iki durumu ayırt edemiyorsa:

$$\forall g \in \mathcal{G} : g(x_1, S_1) = g(x_2, S_2), \quad (89)$$

burada  $\mathcal{G}$ , sistemin davranışsal, bilişsel ve rapora dayalı ölçümler üzerinden uygulayabildiği ayrıştırıcı fonksiyonlar kümesini temsil eder (karar, tepki süresi, sözel rapor, fizyolojik yanıt, vb.).

**Qualia uzayı.** Bu ilişki altında oluşan quotient uzayı,

$$Q = (X \times \mathcal{S}) / \sim \quad (90)$$

qualia uzayı olarak adlandırılır. Her  $[x, S]_{\sim} \in Q$  sınıfı, fonksiyonel olarak ayırt edilemeyen bir fenomenolojik deneyim kümesini temsil eder.

**Koordinatlandırma fonksiyonu.** Pratikte,  $Q$  soyut bir uzay olduğu için, ölçülebilir bir koordinat sistemine ihtiyaç duyulur. Bunun için

$$\Phi : X \times \mathcal{S} \rightarrow \mathbb{R}^m \quad (91)$$

şeklinde bir haritalama tanımlanır; amaç,

$$(x_1, S_1) \sim (x_2, S_2) \iff \Phi(x_1, S_1) = \Phi(x_2, S_2) \quad (92)$$

koşuluna olabildiğince yaklaşmaktır.

**Matematiksel bağlantı.**  $\Phi$  fonksiyonu, quotient space  $Q$ 'nın bir koordinat sistemidir. Tersine mühendislik süreci (5.2, Tahmin 14), bu soyut yapıya ölçülebilir koordinatlar atamayı sağlar. Bu şekilde, hard problem'in ontolojik boyutu askıda bırakılırken, mekanistik boyutu tam olarak operasyonelleştirilmiş olur.

$\Phi$ , aşağıdaki boyutları içeren bir vektör olarak gerçekleştirilebilir:

- fenomenolojik yoğunluk,
- valans/arousal,
- duyuusal modalite bileşenleri,
- zenginlik/yenilik derecesi.

### 3.11.1 Fenomenolojik Bileşenlerin Faktörizasyonu

Yukarıda  $\Phi$ 'nin hangi boyutları içerebileceği genel hatlarıyla belirtilmiştir. Bu alt bölümde, hem fenomenolojik literatürle uyumlu hem de hesaplanabilir bir parametrik  $\Phi$  ailesi öneriyoruz.

**Faktörize yapı.** Qualia vektörünü, fenomenolojik deneyimin farklı yönlerini yakalayan üç ana bileşene ayırıyoruz:

$$\Phi(x, S_t) = (\phi_{\text{modal}}(x, S_t), \phi_{\text{affect}}(x, S_t), \phi_{\text{struct}}(x, S_t)) \in \mathbb{R}^{m_1+m_2+m_3}, \quad (93)$$

Not: Notasyonun sade olması için  $\Phi(x, S_t)$  yazıyoruz; ancak  $\Phi$ , değer fonksiyonu  $U$ , bilinç düzeyi  $L_t$  ve öncelik vektörü  $W_t$  üzerinden dolaylı olarak  $(M_t, W_t)$ 'ye de bağlıdır.

burada:

- $\phi_{\text{modal}}$ : duyuusal modalite kompozisyonu (görsel, işitsel, dokunsal, vb. modalitelerin göreceli baskınlığı),
- $\phi_{\text{affect}}$ : duygulanımsal eksen (valans, arousal ve ilgili duyuusal boyutlar),
- $\phi_{\text{struct}}$ : deneyimin yapısal özellikleri (yoğunluk, yenilik, zamansal dinamik, dikkat keskinliği, be-densellik).

**Modalite Kompozisyonu ( $\phi_{\text{modal}}$ ).** Duyusal durum  $S_t$ 'nin modalitelere ayrılmış alt-uzaylara sahip olduğunu varsayalım (3.3):

$$S_t = (S_t^{(\text{vis})}, S_t^{(\text{aud})}, S_t^{(\text{som})}, S_t^{(\text{int})}, \dots), \quad (94)$$

burada:

- $S_t^{(\text{vis})}$ : görsel bilgi,
- $S_t^{(\text{aud})}$ : işitsel bilgi,
- $S_t^{(\text{som})}$ : somatosensory (dokunma, propriosepsiyon),
- $S_t^{(\text{int})}$ : interoceptive (iç beden durumu: kalp atışı, solunum, visseral sinyaller).

Her modalite için bir aktivasyon enerjisi tanımlarız:

$$e_\ell(S_t) = \|S_t^{(\ell)}\|^2, \quad \ell \in \{\text{vis}, \text{aud}, \text{som}, \text{int}, \dots\}, \quad (95)$$

ve bu enerjiler üzerinden softmax ile normalize edilmiş bir dağılım elde ederiz:

$$p_\ell(S_t) = \frac{\exp(\kappa_m e_\ell(S_t))}{\sum_j \exp(\kappa_m e_j(S_t))}, \quad (96)$$

burada  $\kappa_m > 0$  modalite seçiciliğini kontrol eden bir parametredir (yüksek  $\kappa_m \rightarrow$  daha keskin modalite ayrımı).



Böylece modalite bileşeni:

$$\phi_{\text{modal}}(x, S_t) = (p_{\text{vis}}(S_t), p_{\text{aud}}(S_t), p_{\text{som}}(S_t), p_{\text{int}}(S_t), \dots) \in \Delta^{m_1-1} \quad (97)$$

qualia'nın “hangi modaliteden ne kadar oluştuğu”nu kodlayan bir karışım vektörüdür.

**Fenomenolojik yorum:** Bu bileşen, deneyimin “görsel ağırlıklı mı, işitsel ağırlıklı mı, yoksa bedensel mi” olduğunu temsil eder. Synesthesia gibi çapraz-modal deneyimler, bu vektörde çoklu modalitelerin eşzamanlı yüksek değerler almasıyla karakterize edilir.

**Duygulanımsal Eksen ( $\phi_{\text{affect}}$ ).** Duygulanımsal boyut, fenomenolojinin hedonik ve uyarılma özelliklerini yakalar. Russell’ın circumplex modeli [22] ve duygulanım araştırmalarına uygun olarak, en az iki temel eksen tanımlarız:

**Valans (hedonik değer).** Değer fonksiyonu  $U(x, S_t, M_t) \in \mathbb{R}^k$  ve öncelik vektörü  $W_t \in \Delta^{k-1}$  zaten teoride tanımlıdır (3.4). Valans, bu çok-bileşenli değer yapısından öğrenilebilir bir projeksiyon ile türetilir:

$$\phi_{\text{val}}(x, S_t) = \tanh(v_{\text{val}}^\top U(x, S_t, M_t)), \quad v_{\text{val}} \in \mathbb{R}^k, \quad (98)$$

burada  $v_{\text{val}}$ , pozitif ödül ve negatif ceza eksenlerini seçen öğrenilebilir bir ağırlık vektörüdür.  $\tanh$  fonksiyonu, valansı  $[-1, +1]$  aralığına normalize eder (negatif  $\rightarrow$  olumsuz his, pozitif  $\rightarrow$  olumlu his).

**Arousal (uyarılmışlık).** Arousal, hem seçimin keskinliği hem de genel bilinç düzeyiyle ilişkilidir. Ancak basit bir “yüksek seçim = yüksek arousal” ilişkisi yetersizdir; çünkü panik gibi durumlarda yüksek arousal ama dağınık bilinç olabilir.

Daha kapsamlı bir form öneriyoruz:

$$\phi_{\text{ar}}(x, S_t) = \sigma(\gamma_1 \|U(x, S_t, M_t)\| + \gamma_2 L_t + \gamma_3 W_t^{(\text{threat})}), \quad (99)$$

burada:

- $\|U\|$ : değer vektörünün normu (büyük değerler  $\rightarrow$  yüksek stake durumlar),
- $L_t = D_{\text{KL}}(\pi_t \| P_t)$ : bilinç düzeyi metriği (3.10),
- $W_t^{(\text{threat})}$ : tehdit/tehlike ile ilişkili öncelik bileşeni (varsa),
- $\sigma$ : lojistik fonksiyon,  $[0, 1]$  aralığına normalize eder,
- $\gamma_1, \gamma_2, \gamma_3$ : öğrenilebilir ağırlıklar.

**Fenomenolojik yorum:** Yüksek  $\phi_{\text{ar}}$  değerleri, “uyanık, enerjik, alert” deneyimlere; düşük değerler ise “sakin, durgun, yorgun” deneyimlere karşılık gelir.

**Genişletilmiş duygu boyutları (opsiyonel).** Ek duygusal boyutlar (korku, öfke, merak, üzüntü) için,  $U$  üzerinde benzer projeksiyonlar tanımlanabilir:

$$\phi_{\text{fear}}(x, S_t) = \sigma(v_{\text{fear}}^\top U(x, S_t, M_t)), \quad \phi_{\text{anger}}(x, S_t) = \sigma(v_{\text{anger}}^\top U(x, S_t, M_t)), \quad (100)$$

vb. Bu durumda duygulanımsal bileşen:

$$\phi_{\text{affect}}(x, S_t) = (\phi_{\text{val}}, \phi_{\text{ar}}, \phi_{\text{fear}}, \phi_{\text{anger}}, \dots) \in \mathbb{R}^{m_2}. \quad (101)$$

**Yapısal Özellikler ( $\phi_{\text{struct}}$ ).** Yapısal bileşen, deneyimin “ne kadar yoğun”, “ne kadar yeni”, “ne kadar dinamik” ve “ne kadar odaklanmış” olduğunu yakalar.

**Fenomenolojik yoğunluk.** Deneyimin genel yoğunluğu, iç temsil vektörünün normu ile ölçülebilir:

$$\phi_{\text{int}}(x, S_t) = \rho_1 \|x\|_2, \quad (102)$$

burada  $\rho_1 > 0$  ölçek parametresidir. Yüksek  $\phi_{\text{int}}$ , zengin ve yoğun bir deneyimi; düşük değerler ise hafif ve silik bir deneyimi temsil eder.

**Yenilik / Sürpriz.** Deneyimin ne kadar beklenmedik olduğu, geçmiş deneyimlerle olan uzaklığına bağlıdır. Geçmiş  $T$  adımıdaki  $(x_\tau, S_\tau)$  çiftlerinden öğrenilen bir manifold tahmini  $\mathcal{M}_{\text{past}}$  tanımlayalım (örneğin bir auto-encoder ile). Yenilik:

$$\phi_{\text{nov}}(x, S_t) = \rho_2 d((x, S_t), \mathcal{M}_{\text{past}}), \quad (103)$$

burada  $d(\cdot, \mathcal{M})$  noktanın manifold'a olan uzaklığıdır (örneğin reconstruction error). Yüksek  $\phi_{\text{nov}}$ , beklenmedik, sürpriz deneyimleri yakalar.

**Not:**  $\mathcal{M}_{\text{past}}$ , pratikte son  $T$  adımıdaki örnekler üzerinden öğrenilen bir PCA manifoldu, variational autoencoder latent space'i veya benzeri bir yapı olarak gerçekleştirilebilir.

**Zamansal dinamik.** Deneyimin zamansal değişim hızı:

$$\phi_{\text{temp}}(C_t, S_t) = \rho_3 \|(C_t, S_t) - (C_{t-1}, S_{t-1})\| \quad (104)$$

Hızlı değişen deneyimler (örn. ani korku, sürpriz) yüksek  $\phi_{\text{temp}}$  ile karakterizedir; durağan, sakin deneyimler ise düşük değerler alır.

**Dikkat keskinliği / Odaklanma.** Dikkat odağının ne kadar keskin olduğu, modalite dağılımının entropisi ile ölçülebilir:

$$\phi_{\text{focus}}(x, S_t) = \rho_4 (H_{\text{max}} - H(p_{\text{modal}})), \quad (105)$$

burada  $H(p_{\text{modal}})$ , (96)'teki modalite dağılımının Shannon entropisidir ve  $H_{\text{max}} = \log(m_1)$  maksimum entropidir. Düşük entropi  $\rightarrow$  keskin odak (tek modalite baskın), yüksek entropi  $\rightarrow$  dağınık dikkat (çoklu modaliteler).

**Bedensellik / Interoception.** Eğer  $S_t^{(\text{int})}$  iç beden durumunu içeriyorsa, bu boyutun qualia üzerindeki etkisi:

$$\phi_{\text{body}}(x, S_t) = \rho_5 \|S_t^{(\text{int})}\|, \quad (106)$$

Yüksek  $\phi_{\text{body}}$ , bedene odaklanmış deneyimleri (örn. fiziksel acı, yorgunluk, açlık, nefes farkındalığı) yakalar. Bu, embodied cognition teorileri [23] ile uyumludur.

**Toplu yapısal vektör.** Yapısal bileşen:

$$\phi_{\text{struct}}(x, S_t) = (\phi_{\text{int}}, \phi_{\text{nov}}, \phi_{\text{temp}}, \phi_{\text{focus}}, \phi_{\text{body}}) \in \mathbb{R}^{m_3}. \quad (107)$$

**Neden Bu Faktörizasyon?** Önerilen  $\Phi$  yapısı yalnızca hesaplama kolaylığı için seçilmiş keyfi bir ayrıştırma değildir. Üç temel gerekçeye dayanır:

**(1) Fenomenolojinin tarihsel eksenleri ile biyolojik ayrışmanın hizalanması.** Klasik fenomenoloji (Husserl, Merleau-Ponty), duysal içerik (modalite), duygulanım (valans-arousal) ve deneyimin yapısal özellikleri (odaklanma, yoğunluk, zamansallık) arasında doğal bir ayrım yapar. Nörobilimde de bu üç eksen farklı nöral yollara karşılık gelir: duysal korteksler, limbik sistem ve frontoparietal ağlar. Dolayısıyla seçtiğimiz faktörizasyon hem fenomenolojik hem biyolojik olarak doğal bir bölünmedir.

**(2) Tersine mühendislik yapılabirliği.**  $\Phi$ 'nin parçalanması, öğrenilebilirlik açısından kritik önemdedir. Tek bir dev boyutlu, anlamlandırılmaz bir qualia vektörü yerine:

- modalite,
- duygulanım,
- yapısal nitelikler

gibi ayrı bileşenlere ayrılmış bir yapı, hem FBCT ajanlarından hem insan verisinden öğrenilebilir. Her boyutun açık fenomenolojik karşılığı olduğu için, sinyal-gürültü oranı artar ve  $\Phi$  tersine mühendislik yoluyla daha kolay tahmin edilir. Bu, Hard Problem'i pratik bir rekonstrüksiyon problemine indirger.

**(3) Evrensel temsil gücü ve genelleme.** Bu üç bileşen, hem insan bilincinde hem yapay ajanlarda görülmesi beklenen tüm qualia varyasyonlarını ifade edecek kadar geneldir. Modalite bileşeni duyuşal çeşitliliği, affect bileşeni motivasyonel durumları, yapısal bileşen ise deneyimin geometrisini kapsar. Bu üçlü yapı, farklı bilinç türleri (hayvan, insan, yapay sistemler) arasında ortak karşılaştırma yapılmasını sağlar.

**Sonuç: Seçim keyfi değil, zorunludur.** Bu nedenle, önerilen faktörizasyon:

1. fenomenolojik açıdan doğal,
2. nörobiyolojik açıdan temelli,
3. hesaplamalı açıdan zorunlu,
4. tersine mühendislik açısından optimal

bir seçimdir. Alternatif bir faktörizasyon, ya hesaplamayı zorlaştıracak ya da fenomenolojiyi kaybedecek veya nörobiyolojik karşılıkları zayıflatacaktır.

**Tam  $\Phi$  Fonksiyonu.** Yukarıdaki üç bileşeni birleştirerek:

$$\Phi(x, S_t) = \begin{pmatrix} \phi_{\text{modal}}(x, S_t) \\ \phi_{\text{affect}}(x, S_t) \\ \phi_{\text{struct}}(x, S_t) \end{pmatrix} \in \mathbb{R}^m, \quad m = m_1 + m_2 + m_3 \quad (108)$$

Bu form:

1. **Fenomenolojik olarak zengin:** Russell circumplex modeli, modalite teorisi, embodied cognition gibi klasik yaklaşımlarla uyumlu.
2. **Yorumlanabilir:** Her boyutun net fenomenolojik karşılığı var.
3. **Öğrenilebilir:** Parametreler  $\{v_{\text{val}}, \gamma_i, \rho_i, \kappa_m\}$  Tahmin 14'te açıklanan tersine mühendislik yoluyla, FBÇT ajanlarından veya insan nöral/davranışsal verisinden fit edilebilir.
4. **Modüler:** Yeni boyutlar (örn. sosyal salience, epistemik değer) gerektiğinde eklenebilir.

**Öğrenilebilirlik Stratejisi.** 5.2, Tahmin 14'te detaylandırılan tersine mühendislik sürecinde:

1. FBÇT ajanı eğitilir, her adımda  $(C_t, S_t, Q_t^{\text{self-report}})$  kayıt edilir.
2.  $Q_t^{\text{self-report}}$  vektörü, yukarıda tanımlanan boyutları içerir: modalite kompozisyonu, valans, arousal, yoğunluk, yenilik, vb.
3. Parametreler  $\theta = \{v_{\text{val}}, \gamma_i, \rho_i, \kappa_m, \dots\}$  şu optimizasyon ile öğrenilir:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \|Q_t^{(i)} - \Phi_{\theta}(C_t^{(i)}, S_t^{(i)})\|^2. \quad (109)$$

4. Öğrenilen  $\hat{\Phi}_{\theta}$ , yeni durumlarda test edilir ve cross-agent tutarlılığı değerlendirilir.

**Fenomenolojik Literatürle Bağlantılar.** Bu  $\Phi$  yapısı, çeşitli fenomenolojik ve psikolojik teorilerle doğal bağlantılara sahiptir:

- **Russell'in Circumplex Modeli** [22]:  $(\phi_{\text{val}}, \phi_{\text{ar}})$  eksenleri doğrudan bu modelle uyumlu.
- **Embodied Cognition** [23]:  $\phi_{\text{body}}$  bileşeni, bedensel temellendirilmiş bilinç teorileriyle örtüşür.
- **Predictive Processing** [16]:  $\phi_{\text{nov}}$  (yenilik/sürpriz) tahmin hatası kavramıyla doğrudan ilişkilidir.
- **Nörofenomenoloji** [24]:  $\Phi$ 'nin parametrelerinin nöral aktiviteden öğrenilmesi, birinci şahıs raporlar ile üçüncü şahıs ölçümler arasında köprü kurar.

Bu detaylı yapılandırma,  $\Phi$ 'nin tamamen soyut bir "kara kutu" olmaktan çıkıp, fenomenolojik olarak anlamlı, matematiksel olarak kesin ve deneysel olarak öğrenilebilir bir fonksiyon ailesi haline geldiğini gösterir. Hard problem, böylece ontolojik bir muammadan, *öğrenilebilir bir projeksiyon problemine* indirgenir.

Not: Burada verilen  $\Phi$  yapısı zorunlu değil; fenomenolojik boyutları yakalayan örnek bir parametrik formdur. Farklı araştırma programları farklı  $\Phi$  parametrizasyonları kullanabilir.

**Qualia Mapping Sınırlamaları  $\Phi$ .** Önerdiğimiz  $\Phi$  fonksiyonu yorumlanabilir ve öğrenilebilir olmasına rağmen tam bir fenomenolojik izomorfizm değildir. Aşağıdaki sınırlılıkları kabul ediyoruz:

1.  $\Phi$  **eksik** olabilir: tüm fenomenolojik boyutları yakalamak zorunda değildir.
2.  $\Phi$  **canlı türüne göre değişebilir**: her tür kendi qualia uzayının farklı projeksiyonlarına sahip olabilir.
3.  $\Phi$  yalnızca *fonksiyonel ayırt edilemezlik* sınıflarını temsil eder — fenomenolojik özün ontolojik açıklamasını vermez.
4.  $\Phi$  tam bir izomorfizm değil, yüksek boyutlu fenomenolojik uzayın *hesaplanabilir bir projeksiyonudur*.

Bu sınırlılıklar,  $\Phi$ 'nin teoriyi zayıflatmaz; aksine, hard problem'i ontolojik bir muammadan *ölçülebilir ve test edilebilir bir projeksiyon problemine* indirger. Teorinin başarısı,  $\Phi$ 'nin "mutlak doğru" olup olmadığıyla değil, ne ölçüde *tutarlı, öğrenilebilir ve tahmin edici* olduğuyla ölçülmelidir.

### 3.12 Fenomenal Durum Uzayı ve Fonksiyonel Eşleme

#### 3.12.1 Motivasyon ve kapsam

Bilinç teorilerinin en tartışmalı yönlerinden biri, fonksiyonel/ hesaplamalı süreçlerin öznel deneyimle nasıl ilişkilendiği sorusudur (hard problem of consciousness). FBÇT bu soruya *ontolojik* bir cevap sunmaya çalışmaz; bunun yerine daha mütevazı, fakat deneysel olarak test edilebilir bir hedef güder:

**Hedef:** Hangi fonksiyonel durumların benzer öznel raporlara, davranışsal seçimlere ve fenomenolojik yargılara yol açması gerektiğini belirleyen bir eşleme şeması tanımlamak.

Bu yaklaşım, “neden öznel deneyim var?” sorusunu askıya alır ve bunun yerine “hangi fonksiyonel durumlar, öznenin ifadesine göre fenomenal olarak benzer/different sayılmalıdır?” gibi operasyonel sorulara odaklanır. Böylece FBÇT, fonksiyonel durumlar ile fenomenolojik benzerlik yargıları arasında yapı-koruyan bir köprü kurmayı hedefler.

#### 3.12.2 Fonksiyonel durum vektörü

FBÇT'nin önceki bölümlerinde tanımlanan temel değişkenler kullanılarak, zaman  $t$ 'deki fonksiyonel durum, sabit boyutlu bir vektör olarak özetlenebilir:

$$z_t = \begin{bmatrix} \phi_S(S_t) \\ \phi_M(M_t) \\ W_t \\ C_t \\ \beta_t \\ \alpha_t \\ \kappa_t \end{bmatrix} \in \mathbb{R}^D,$$

burada:

- $\phi_S(S_t) \in \mathbb{R}^{d_S}$ : Duyusal durumun (Bölüm 3.4.1) kodlanmış temsili,
- $\phi_M(M_t) \in \mathbb{R}^{d_M}$ : Bellek durumunun özet temsili,
- $W_t \in \mathbb{R}^k$ : Öncelik/değer vektörü,
- $C_t \in \mathbb{R}^{d_C}$ : Çöküş operatörü tarafından seçilen bilinçli içerik,
- $\beta_t \in \mathbb{R}$ : Sıcaklık parametresi (Bölüm 3.1.6),
- $\alpha_t \in \Delta^2$ : Bileşen ağırlıkları  $(\alpha_{S,t}, \alpha_{M,t}, \alpha_{W,t})$  (Bölüm 3.6),
- $\kappa_t \in \mathbb{R}^{d_\kappa}$ : Bedensel durum, duygusal ton ve dikkat gibi ek bağlamsal değişkenlerin özeti.

Toplam boyut  $D = d_S + d_M + k + d_C + 1 + 3 + d_\kappa$  olup pratik uygulamalarda tipik olarak  $10^2$ – $10^3$  aralığında seçilebilir. Böylece  $\mathcal{Z} \subset \mathbb{R}^D$ , FBÇT ajanının fonksiyonel durum uzayı olarak düşünülebilir.

### 3.12.3 Fenomenal durum uzayı

Fenomenolojik deneyimlerin temsil edildiği soyut bir uzay tanımlarız:

$$\mathcal{Q} = (\mathcal{Q}, d_{\mathcal{Q}}),$$

burada  $\mathcal{Q}$ , fenomenal durumların kümesi;  $d_{\mathcal{Q}}$  ise bu durumlar arasındaki öznel benzerliği nicelleştiren bir metriktir. Sezgisel olarak:

$$d_{\mathcal{Q}}(q_1, q_2) \text{ küçük} \Rightarrow q_1 \text{ ve } q_2 \text{ deneyim olarak birbirine çok benzer,}$$

$$d_{\mathcal{Q}}(q_1, q_2) \text{ büyük} \Rightarrow \text{deneyimler fenomenolojik olarak belirgin biçimde farklıdır.}$$

Teorik düzeyde  $\mathcal{Q}$  yalnızca bir metrik uzay olarak varsayılır; pratikte ise genellikle  $\mathcal{Q} \subset \mathbb{R}^{d_{\mathcal{Q}}}$  olarak öğrenilmiş bir gömülü uzay (embedding) biçiminde alınabilir.

### 3.12.4 Fonksiyonel durumdan fenomenal duruma eşleme: $\Phi$

Fonksiyonel durumdan fenomenal duruma

$$\Phi : \mathcal{Z} \rightarrow \mathcal{Q}, \quad q_t = \Phi(z_t),$$

eşleme fonksiyonu ile geçilir. FBC'T açısından bilinçli deneyimin *çekirdek taşıyıcısı* çöküş sonucu seçilen içerik  $C_t$ 'dir; ancak deneyimin tam fenomenal profili, bağlamı oluşturan diğer bileşenlere de bağlıdır:

$$q_t = \Phi(\phi_S(S_t), \phi_M(M_t), W_t, C_t, \beta_t, \alpha_t, \kappa_t).$$

Fonksiyonel ve fenomenal benzerlik arasında zayıf bir süreklilik varsayımı yaparız.  $\mathcal{Z}$  üzerinde, bileşenlerin ölçeklerine uygun bir metrik  $d_{\mathcal{Z}}$  seçildiğinde,  $\Phi$  için Lipschitz türü bir kısıt öne sürülür:

$$d_{\mathcal{Q}}(\Phi(z_1), \Phi(z_2)) \leq L d_{\mathcal{Z}}(z_1, z_2),$$

bir  $L > 0$  sabiti için. Bu, fonksiyonel durumdaki küçük değişimlerin, fenomenolojik uzayda keyfi derecede büyük sıçramalara yol açmaması gerektiğini ifade eder; bu da fenomenolojik sezgiler ve klinik gözlemlerle uyumludur.

### 3.12.5 Eşdeğerlik sınıfları ve “aynı fenomenal durum“

Öznenin ayırt edemediği küçük farkları modellemek için,  $\mathcal{Q}$  üzerinde bir eşdeğerlik ilişkisi tanımlayabiliriz. Bir  $\varepsilon > 0$  eşiği verildiğinde:

$$q_1 \sim q_2 \iff d_{\mathcal{Q}}(q_1, q_2) < \varepsilon.$$

Bu durumda her eşdeğerlik sınıfı

$$[q] = \{q' \in \mathcal{Q} \mid q' \sim q\}$$

öznenin fenomenolojik olarak “aynı” kabul ettiği bir durum kümesini temsil eder. Böylece “aynı kırmızıyı görüyorum” gibi ifadeler, bu eşdeğerlik sınıfları üzerinden formel hale getirilebilir.

### 3.12.6 Deneyisel veri üzerinden $\Phi$ 'nin öğrenilmesi

Teoride  $\Phi$  doğrudan tanımlansa da, fenomenal duruma doğrudan erişim mümkün değildir; yalnızca raporlar, benzerlik yargıları ve seçim davranışları gözlemlenebilir. Bu nedenle  $\Phi$ , dolaylı veri üzerinden öğrenilen parametrik bir fonksiyon olarak ele alınır.

Deneyisel bir paradigmayı düşünelim: her denemede sistem bir fonksiyonel durum  $z_i$  üretir ve özne, bu duruma ilişkin fenomenolojik raporu  $r_i$  verir (örneğin bir skala, etiket veya benzerlik yargısı). Özellikle üçlü karşılaştırmalar (triplet) şu tür kısıtlar üretir:

$$“z_i \text{ deneyimi, } z_j \text{'ye, } z_k \text{'ye olduğundan daha çok benziyor.”}$$

Bu, aşağıdaki sıralama kısıtına dönüştürülebilir:

$$d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_j)) + \delta \leq d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_k)),$$

burada  $\delta > 0$  bir marjdır. Tüm bu kısıtları minimize eden bir kayıp fonksiyonu:

$$\mathcal{L}_{\Phi} = \sum_{(i,j,k) \in \mathcal{T}} \left[ \delta + d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_j)) - d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_k)) \right]_+,$$

üzerinden  $\Phi$  parametrik olarak (örneğin bir sinir ağı) öğrenilebilir. Sonuçta ortaya çıkan  $\mathcal{Q}$  ve  $d_{\mathcal{Q}}$ , fenomenolojik benzerlik yargılarına en iyi uyan gömülü uzay olarak yorumlanır.

### 3.12.7 Somut örnek: renk ve duygusal ton

Renk algısı, fenomenal gömülü uzayın yapısını somutlaştırmak için klasik bir örnektir. Basitleştirilmiş bir durumda,  $C_t$  “bir yüzeyin yansıttığı dalga boyu dağılımını temsil eden” bir durum olsun ve  $W_t$  bu uyaranla ilişkili duygusal/motivasyonel bileşenleri içersin.

Deneyssel olarak, öznel renk benzerliklerine dayalı bir gömme ile üç boyutlu bir renk uzayı elde edilebilir; FBÇT perspektifinden bu,  $\mathcal{Q}$ ’nın renk alt uzayının yaklaşık bir örneğidir. Bu durumda

$$q_t^{(\text{renk})} = \Phi_{\text{renk}}(\phi_S(S_t), C_t, \kappa_t) \in \mathbb{R}^3$$

olarak düşünülebilir. Aynı fiziksel uyaran (aynı  $C_t$ ) farklı bağlamlarda (farklı  $W_t$  ve  $\kappa_t$ ) algısal olarak aynı renge karşılık gelebilir, ancak  $\mathcal{Q}$  içinde renk alt uzayına eklenen duygusal boyutlarda farklı noktalara haritalanabilir. Böylece “*aynı rengi hem çekici hem rahatsız edici bulma*” gibi olgular, aynı renk alt-manifold’unda farklı duygusal koordinatlara sahip fenomenal durumlar olarak modellenenebilir.

### 3.12.8 Deneyssel test edilebilirlik

Fonksiyonel-fenomenal eşleme  $\Phi$  için FBÇT, aşağıdaki türde test edilebilir tahminler üretir:

**Test 1: Benzerlik yargılarının öngörüsü.** İki fonksiyonel durum  $z_i, z_j$  verildiğinde model:

$$d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_j)) \propto \text{öznel benzerlik derecesi}$$

öngörüsünde bulunur. Deneyssel paradigmalarda, farklı uyaran kombinasyonlarının öznel benzerlik puanları ile model tarafından hesaplanan  $d_{\mathcal{Q}}$  arasındaki korelasyon, eşlemenin geçerliliği için bir ölçüt sunar.

**Test 2: Çapraz-modal benzerlik.** Görsel ve işitsel uyaranlar gibi farklı modalitelere ait durumların, aynı fenomenal alt-manifold’a projekte edilip edilmediği test edilebilir. Örn. yüksek frekanslı bir ses ile yüksek parlaklıklı bir ışık, benzer “uyanıklık/arousal” boyutuna map oluyorsa  $d_{\mathcal{Q}}$  küçük olmalıdır.

**Test 3: Klinik disosiasyon.** Depersonalizasyon veya derealizasyon gibi durumlarda, fonksiyonel bileşenler (örneğin  $\phi_S(S_t)$  ve  $C_t$ ) görece korunmuş olabilir; ancak öznel rapor “gerçek dışılık” hissi barındırır. FBÇT perspektifinden bu, fonksiyonel durum  $z_t$ ’nin normal aralıkta olmasına karşın  $\Phi$ ’nin çıktısının, sağlıklı deneklerin  $\mathcal{Q}$ ’sine göre marjinal bölgelerde yer almasıyla modellenenebilir; bu da öğrenilmiş  $\Phi$ ’ye göre yüksek “residual error” ile ölçülebilir.

Bu tür testler, Bölüm 6.2’te önerilen nöral tahminlerle birlikte, FBÇT’nin fenomenal eşleme bileşenini falsifiye edilebilir bir çerçeve haline getirir.

### 3.12.9 Nörobiyolojik substrat

Fonksiyonel durum  $z_t$  ile nöral durum  $B_t$  arasındaki ilişkiyi, iki aşamalı bir şema ile düşünebiliriz:

$$B_t \xrightarrow{\rho} z_t \xrightarrow{\Phi} q_t.$$

Burada  $\rho$ , fiziksel beyin durumundan (örneğin geniş ölçekli aktivite desenlerinden) FBÇT’nin fonksiyonel bileşenlerine giden haritalamayı temsil eder. Tablo 4,  $z_t$ ’nin bileşenleri için olası nöral karşılıkları özetler.

Bu tablo, FBÇT’nin fenomenal eşleme bileşeninin, özgül nöral ölçümler üzerinden dolaylı olarak test edilmesine olanak tanır: benzer  $z_t$  ve dolayısıyla benzer  $q_t$  beklenen durumlarda, ilgili bölgelerde benzer aktivasyon desenleri gözlenmelidir.

### 3.12.10 Kapsam ve sınırlamalar

Bu formalizm *şunları yapmaz*:

- Qualia’nın ontolojik doğasını açıklamaz,
- “Neden fenomenal deneyim var?” sorusuna cevap vermez,
- Hard problem’i çözmeye iddiasında bulunmaz,
- Zombi veya Mary’s Room gibi düşünce deneylerini doğrudan çürütmeye çalışmaz.

Buna karşılık, *şunları yapar*:

$z_t$ bileşeni	Olası nöral substrat	Ölçülebilir gösterge
$\phi_S(S_t)$	V1–V4, birincil/sekonder duyuşal korteks	BOLD sinyali, lokal alan potansiyeli
$\phi_M(M_t)$	Hippocampus, medial temporal lob	Theta gücü, pattern completion imzaları
$W_t$	OFC, vmPFC, VTA	fMRI aktivitesi, dopaminerjik sinyaller
$C_t$	Frontoparietal ağ	GNW-benzeri geniş ölçekli aktivasyon
$\beta_t$	ACC, LC (arousal sistemleri)	Pupilla genişliği, EEG uyarılma indeksleri
$\alpha_t$	DLPFC, bağlam-duyarlı prefrontal ağlar	Görev bağlamına bağlı modülasyon desenleri
$\kappa_t$	Insula, somatotopik alanlar	Enteroseptif ve bedensel durum göstergeleri

Table 4: Fonksiyonel durum vektörünün bileşenleri için olası nöral eşlemeler.

- Fonksiyonel eşdeğerlik için, fenomenal benzerlik bazlı operasyonel kriterler sunar,
- Öznel raporları ve benzerlik yargılarını, sistemin fonksiyonel değişkenleriyle ilişkilendirir,
- Deneyisel olarak test edilebilir tahminler üretir (benzerlik yargıları, çapraz-modal benzerlik, klinik dissosiasyon),
- Farklı bilişsel mimarilerin, aynı fenomenal uzaya gömülüp gömülemeyeceğini incelemek için ortak bir çerçeve sağlar.

FBÇT'nin iddiası, kısaca, *aynı fonksiyonel durum vektörünün (zaman ve gürültü farkları dışında) aynı davranışsal/fenomenal profile yol açması gerektiğidir*. Bu, metafizik tartışmaları askıya alarak, bilinç araştırmalarını deneyisel olarak erişilebilir bir seviyeye çekmeyi amaçlar.

### 3.13 Algoritma: FBÇT Ajamı

Teorinin hesaplamalı uygulamaları ve yapay sistemlerdeki testleri için, tam FBÇT dinamiğini uygulayan bir algoritma sunulur.

---

**Algorithm 1** FBÇT Ajanı – Tek Zaman Adımı

---

**Require:**  $S_t$  (duyusal girdi),  $M_t$  (bellek),  $W_t$  (ağırlık),  $\beta_t$  (sıcaklık)

**Ensure:**  $C_t$  (bilinç durumu),  $M_{t+1}$ ,  $W_{t+1}$

```
1: Aday Üretimi:
2:    $\mu_t \leftarrow G(S_t, M_t, W_t)$ 
3:   Örnekler:  $\{x_t^{(i)}\}_{i=1}^N \sim \mu_t$ 
4: Değer Hesaplama:
5: for  $i = 1, \dots, N$  do
6:    $U^{(i)} \leftarrow U(x_t^{(i)}, S_t, M_t)$ 
7:    $f^{(i)} \leftarrow \langle W_t, U^{(i)} \rangle$ 
8:   if self-model aktif then
9:      $f^{(i)} \leftarrow f^{(i)} - \lambda \|x_t^{(i), \text{self}} - F_{\text{self}}(H_t)\|^2$ 
10:  end if
11: end for
12: Politika Dağılımı:
13:    $\pi_t^{(i)} \leftarrow \frac{\exp(\beta_t f^{(i)})}{\sum_{j=1}^N \exp(\beta_t f^{(j)})}$ 
14: Çöküş:
15:    $k \sim \text{Categorical}(\pi_t)$ 
16:    $C_t \leftarrow x_t^{(k)}$ 
17: Geri Bildirim:
18:   Gözlem:  $R_t$  (ödül/ceza sinyali)
19:    $\Delta M_t \leftarrow \eta_M \cdot h(C_t, S_t, M_t)$ 
20:    $M_{t+1} \leftarrow M_t + \Delta M_t$ 
21:    $\Delta W_t \leftarrow \alpha_W \cdot g(R_t, W_t, C_t)$ 
22:    $W_{t+1} \leftarrow \Pi_{\Delta^{k-1}}(W_t + \Delta W_t)$ 
   return  $C_t, M_{t+1}, W_{t+1}$ 
```

---

Bu algoritma, simülasyon çalışmalarında (6.4, Tahmin 7-8) ve yapay bilinç adaylarının test edilmesinde (6.5, Tahmin 9-10) doğrudan kullanılabilir.

### 3.14 Teorinin Derli Toplu Özeti

FBÇT, bilinç dinamiğini aşağıdaki temel yapı taşları üzerinden formelleştirir:

Durum uzayları:	$S_t \in \mathbb{R}^{n_S}, M_t \in \mathbb{R}^{n_M}, W_t \in \Delta^{k-1}, X_t \in \mathbb{R}^{n_X}$
Aday üretimi:	$\mu_t(dx) = G(S_t, M_t, W_t)(dx)$
Değer fonksiyonu:	$U : X \times \mathcal{S} \times \mathcal{M} \rightarrow \mathbb{R}^k, f = \langle W, U \rangle$
Politika ve çöküş:	$\pi_t(dx) \propto \exp(\beta_t f(x, S_t, M_t, W_t)) \mu_t(dx), C_t \sim \pi_t$
Geri bildirim:	$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t)$
Benlik modeli:	$\ x_t^{\text{self}} - F_{\text{self}}(H_t)\  \leq \varepsilon$
Bilinç düzeyi:	$L_t = D_{\text{KL}}(\pi_t \  P_t)$
Qualia yapısı:	$Q = (X \times \mathcal{S}) / \sim, \Phi : X \times \mathcal{S} \rightarrow \mathbb{R}^m$

(110)

Bu çerçeve, bilinci

- hesaplanabilir bir seçim süreci,
- dinamik ve geri beslemeli bir kontrol sistemi,
- ve fenomenolojik içeriği quotient uzayında yapılandırılmış bir durum

olarak ele alır. Sonraki bölümlerde, bu formel yapının klasik felsefi sorulara nasıl sistematik cevaplar ürettiği (5) ve hangi deneysel tahminleri gerektirdiği (6) detaylandırılacaktır.



## 4 Temel İlkeler ve Varsayımlar

Önceki bölümlerde FBÇT'nin matematiksel yapısını geliştirdik: çöküş operatörü  $C_t$ , jeneratif süreç  $G$ , değer fonksiyonu  $f$ , ve fenomenal eşleme  $\Phi$ . Bu bölümde teoremin dayandığı temel ilkeleri açıkça ifade ediyoruz. Bu ilkeler hem FBÇT'nin yapısal özelliklerini özetler hem de test edilebilir tahminler için temel oluşturur.

### 4.1 İlke 1: Fonksiyonel Durum Tamlığı

**İfade.** Zaman  $t$ 'deki fonksiyonel durum vektörü

$$z_t = (\phi_S(S_t), \phi_M(M_t), W_t, C_t, \beta_t, \alpha_t, \kappa_t) \in \mathbb{R}^D,$$

öznenin bilinçli deneyimi hakkındaki tüm fonksiyonel olarak ilgili bilgiyi içerir.

**Sezgi.** İki sistem aynı  $z_t$  durumuna sahipse, fonksiyonel açıdan ayırt edilemezler ve aynı fenomenal profile yol açmalıdırlar. Bu, bilinç için "fonksiyonel yeterlilik" ilkesidir.

**Sonuç.** Fenomenal benzerlik yargıları, yalnızca  $z_t$  bileşenlerinin benzerliğine bağlıdır. Mikro-fiziksel implementasyon detayları (multiple realizability) fenomenal durumu değiştirmez.

### 4.2 İlke 2: Fenomenal Süreklilik

**İfade.** Fonksiyonel durum uzayı  $\mathcal{Z}$  ve fenomenal durum uzayı  $\mathcal{Q}$  arasındaki eşleme  $\Phi : \mathcal{Z} \rightarrow \mathcal{Q}$  Lipschitz süreklidir:

$$d_{\mathcal{Q}}(\Phi(z_1), \Phi(z_2)) \leq L d_{\mathcal{Z}}(z_1, z_2),$$

burada  $L > 0$  bir Lipschitz sabitidir.

**Sezgi.** Fonksiyonel durumda küçük değişimler, fenomenal deneyimde keyfi derecede büyük sıçramalara yol açamaz. Beyin durumundaki kademeli değişimler, deneyimde de kademeli değişimlere karşılık gelir.

**Sonuç.** Bu kısıt,  $\Phi$ 'nin benzerlik yargıları üzerinden öğrenilebilir olmasını sağlar ve "fenomenal uçurum" (ani sıçramalar) senaryolarını dışlar. Klinik olarak, beyin hasarının kademeli fenomenal değişimlere yol açması beklenir.

### 4.3 İlke 3: Çöküş-Temelli İçerik Belirlenimi

**İfade.** Bilinçli içeriğin çekirdeği, çöküş operatörü tarafından seçilen durumdur:

$$C_t \sim \pi_t(\cdot \mid S_t, M_t, W_t),$$

burada  $\pi_t$  Boltzmann-softmax dağılımıdır. Ancak tam fenomenal durum yalnızca  $C_t$ 'den ibaret değildir:

$$q_t = \Phi(z_t) = \Phi(\phi_S(S_t), \phi_M(M_t), W_t, C_t, \beta_t, \alpha_t, \kappa_t).$$

**Sezgi.** Bilinç hem *ne* seçildiğine ( $C_t$ ) hem de *hangi bağlamda* seçildiğine ( $W_t, \beta_t, \alpha_t, \kappa_t$ ) bağlıdır. Aynı içerik farklı duygusal/motivasyonel bağlamlarda farklı fenomenal deneyimlere yol açar.

**Sonuç.** Bu, "aynı algısal içeriği hem tehdit hem de masum görebilme" gibi bağlam-bağımlı fenomenolojik varyasyonları açıklar (örn. bahçe senaryosu, Bölüm 3.5.6).

### 4.4 İlke 4: Fonksiyonel Ayırt Edilemezlik

**İfade.** Eğer iki fonksiyonel durum bir eşik  $\varepsilon_{\text{func}} > 0$  içinde yakınsa:

$$d_{\mathcal{Z}}(z_1, z_2) < \varepsilon_{\text{func}},$$

o zaman fenomenal olarak aynı eşdeğerlik sınıfına aittirler:

$$[\Phi(z_1)] = [\Phi(z_2)].$$

**Sezgi.** Çok küçük fonksiyonel farklar öznel olarak ayırt edilemez. Bu, “tam aynı kırmızıyı gördüm” gibi fenomenolojik eşdeğerlik yargılarının temelini verir.

**Sonuç.** Fenomenal durum uzayı  $\mathcal{Q}$ , sürekli bir manifold olsa da, pratik ölçümler sonlu sayıda ayırt edilebilir sınıfa karşılık gelir.

#### 4.5 İlke 5: Fenomenal Eşlemenin Öğrenilebilirliği

**İfade.** Eşleme fonksiyonu  $\Phi : \mathcal{Z} \rightarrow \mathcal{Q}$  soyut bir postüla değildir; deneysel veri üzerinden öğrenilebilir parametrik bir fonksiyon ailesidir. Triplet karşılaştırmaları, benzerlik yargıları ve davranışsal seçimler  $\Phi$  için kısıtlar üretir (Bölüm 3.6.6).

**Sezgi.**  $\Phi$ , fenomenal benzerlik yargılarına en iyi uyan gömülü uzay olarak ortaya çıkar. Bu, FBÇT’yi deneysel olarak test edilebilir kılar.

**Sonuç.**  $\Phi$  öğrenildikten sonra, yeni fonksiyonel durumlar için fenomenal tahminler yapılabilir. Klinik olarak, “bu beyin durumu hangi deneyime karşılık gelir?” sorusu cevaplanabilir hale gelir.

#### 4.6 İlke 6: Nöral Alt-Temsil

**İfade.** Fonksiyonel durum vektörünün her bileşeni

$$(\phi_S, \phi_M, W_t, C_t, \beta_t, \alpha_t, \kappa_t),$$

özgül nöral yapılar ve ölçülebilir nöral göstergelerle ilişkilendirilebilir (Tablo 3.1). Benzer  $z_t$  durumları, ilgili beyin bölgelerinde benzer aktivasyon desenlerine yol açar.

**Sezgi.** FBÇT’nin fonksiyonel bileşenleri tamamen soyut değildir; beynin fiziksel yapısına “ground” edilir. Her bileşen için nöral karşılık önerilebilir ve test edilebilir.

**Sonuç.** Bu ilke, FBÇT’yi nörogörüntüleme ve elektrofizyoloji ile test edilebilir hale getirir. Örneğin,  $\beta_t$  artışının ACC ve LC aktivasyonu ile korelasyon göstermesi beklenir (Bölüm 3.1.6).

#### 4.7 İlke 7: Bağlamsal Duyarlılık

**İfade.** Aynı çöküş içeriği  $C_t$ , farklı bağlamsal durumlarda  $(W_t, \beta_t, \alpha_t, \kappa_t)$  farklı fenomenal durumlara yol açabilir:

$$\Phi(\dots, C_t, W_t^{(1)}, \dots) \neq \Phi(\dots, C_t, W_t^{(2)}, \dots).$$

**Sezgi.** Deneyim yalnızca “ne görüldüğüne” değil, aynı zamanda duygu, motivasyon, uyarılma ve dikkat gibi bağlamsal faktörlere de bağlıdır. Aynı duyuşsal input, farklı psikolojik durumlarda radikal biçimde farklı deneyimlere yol açabilir.

**Sonuç.** Bu, tehdit algısı, duygusal renklenme ve dikkat modülasyonunun fenomenolojik etkilerini açıklar. Örneğin, yüksek  $\alpha_{W,t}$  altında “gölge”  $\rightarrow$  “tehdit” olarak deneyimlenir; düşük  $\alpha_{W,t}$  altında ise “belirsiz nesne” olarak kalır (Bölüm 3.3.1).

#### 4.8 İlke 8: Fenomenal Eşdeğerlik Sınıfları

**İfade.** Fenomenal durum uzayı  $\mathcal{Q}$  üzerinde bir eşik  $\varepsilon > 0$  için eşdeğerlik ilişkisi tanımlanır:

$$q_1 \sim q_2 \iff d_{\mathcal{Q}}(q_1, q_2) < \varepsilon.$$

Her eşdeğerlik sınıfı  $[q]$ , öznenin fenomenolojik olarak “aynı” kabul ettiği durumlar kümesidir.

**Sezgi.** Öznel deneyimde sürekli varyasyon olsa da, pratik olarak sonlu sayıda ayırt edilebilir kategori vardır. “Aynı kırmızı” veya “aynı acı” gibi ifadeler bu eşdeğerlik sınıflarına karşılık gelir.

**Sonuç.** Bu yapı, fenomenal raporların neden kategorik olduğunu (örn. renk isimleri, acı skalaları) ve benzerlik yargılarının nasıl tutarlı olabildiğini açıklar.

#### 4.9 İlke 9: Çoklu Gerçeklenebilirlik

**İfade.** Aynı fonksiyonel durum  $z_t$ , farklı mikro-fiziksel substratlar üzerinde gerçekleştirilebilir:

$$B_t^{(1)} \xrightarrow{\rho} z_t \xleftarrow{\rho} B_t^{(2)},$$

burada  $B_t$  fiziksel beyin durumu ve  $\rho$  nöral-fonksiyonel haritalamadır. Her iki substrat da aynı fenomenal duruma yol açar:

$$\Phi(z_t) = q_t.$$

**Sezgisel açıklama.** Bilinç, belirli nöronların veya moleküllerin varlığına değil, fonksiyonel organizasyona bağlıdır. Aynı hesaplamalı yapı farklı fiziksel sistemlerde (biyolojik, silikon, hibrit) gerçekleştirilebilir.

**Sonuç.** Bu ilke, FBÇT'nin türler arası bilinç karşılaştırmalarına ve yapay bilinç tartışmalarına olanak tanır. Farklı türler (insan, hayvan, yapay sistem) aynı  $\mathcal{Q}$  uzayına gömülebilirse, fenomenal karşılaştırmalar yapılabilir.

#### 4.10 İlke 10: Test Edilebilirlik

**İfade.** FBÇT'nin her yapısal bileşeni için somut, falsifiye edilebilir tahminler üretilebilir:

- Benzerlik yargıları:  $d_{\mathcal{Q}}(\Phi(z_i), \Phi(z_j))$  öznel puanlarla korelasyon göstermelidir ( $r^2 > 0.7$ , Bölüm 3.6.8).
- Nöral karşılıklar:  $z_t$  bileşenleri belirtilen beyin bölgelerinde ölçülebilir aktivasyonla eşleşmelidir (Tablo 3.1).
- Çöküş mekanizması:  $\beta_t$  manipülasyonu (örn. stres, arousal) seçim dağılımını öngörülebilir biçimde değiştirmelidir.
- Bağlamsal modülasyon:  $\alpha_t$  değişimleri, aynı duyuşal girdiden farklı fenomenal raporlara yol açmalıdır.

**Sezgi.** FBÇT yalnızca matematiksel bir çerçeve değil, deneysel paradigmalara test edilebilen bilimsel bir teodir. Her ilke somut deneylere karşılık gelir.

**Sonuç.** Bu ilke, FBÇT'yi metafizik spekülasyondan ayırır ve bilişsel bilim ile nörobilimin standart yöntemleriyle değerlendirilebilir kılar.

#### 4.11 İlkelerin Karşılıklı İlişkileri

Yukarıdaki on ilke birbirinden bağımsız değildir; aksine FBÇT'nin tutarlı bir yapı oluşturmaları için birlikte çalışırlar:

- **İlke 1 ve 9:** Fonksiyonel tamlık + çoklu gerçekleştirilebilirlik  $\rightarrow$  substrate-independent consciousness.
- **İlke 2 ve 5:** Fenomenal süreklilik + öğrenilebilirlik  $\rightarrow \Phi$  empirically constrainable.
- **İlke 3 ve 7:** Çöküş-temelli içerik + bağlamsal duyarlılık  $\rightarrow$  same input, different experience.
- **İlke 6 ve 10:** Nöral substrat + test edilebilirlik  $\rightarrow$  neuroscientifically grounded predictions.

Bu ilişkiler, FBÇT'nin yalnızca bir bileşen koleksiyonu değil, entegre bir sistem olduğunu gösterir.

#### 4.12 Sınırlamalar ve Kapsam Dışı Kalanlar

Bu ilkeler FBÇT'nin neyi *açıkladığını* tanımlarken, neyi *açıklamadığını* da netleştirir:

### Açıklamaz:

- **Hard problem:** Neden fenomenal deneyim var? (ontolojik soru)
- **Qualia'nın doğası:** "Kırmızılık nedir?" (metafizik soru)
- **Zombi senaryoları:** Fonksiyonel eşdeğer ama fenomenal olarak farklı sistemler mümkün mü? (düşünce deneyi)

### Açıklar:

- **Fonksiyonel eşdeğerlik:** Hangi sistemler aynı fenomenal profile sahiptir?
- **Fenomenal benzerlik:** Hangi deneyimler birbirine daha yakındır?
- **Bağlamsal modülasyon:** Aynı input nasıl farklı deneyimlere yol açar?
- **Nöral karşılıklar:** Hangi beyin durumları hangi deneyimlerle ilişkilidir?

Bu ayırım, FBÇT'nin bilimsel olarak test edilebilir kalırken aşırı iddiacı olmaktan kaçınmasını sağlar.

## 4.13 Özet

FBÇT'nin on temel ilkesi, teorinin:

1. **Fonksiyonel temelini** (İlke 1, 4, 9),
2. **Fenomenal yapısını** (İlke 2, 3, 8),
3. **Deneyisel bağlantısını** (İlke 5, 6, 10),
4. **Açıklayıcı kapsamını** (İlke 7)

tanımlar.

Bu ilkeler birlikte, FBÇT'yi hem matematiksel olarak titiz hem de deneyisel olarak test edilebilir bir bilinç teorisi haline getirir. Sonraki bölümlerde (Bölüm 6), bu ilkelerden türetilen somut deneyisel tahminleri sunacağız.

## 5 Klasik Felsefi Problemlere Çözümler

FBÇT'nin ayırt edici özelliği, klasik felsefi problemleri "ekstra metafizik varsayımlar" olmadan, doğrudan *fonksiyonel* ve *matematiksel* düzeyde yeniden formüle etmesidir. Teori, bilinci açıklamak için yeni bir madde türü, gizemli bir "iç özne" veya fizik-dışı nedenlere başvurmaz; bunun yerine, daha önceki bölümlerde tanımlanan durum uzayları

$$S_t, M_t, W_t \in \mathcal{S}, \mathcal{M}, \mathcal{W}, \quad C_t \in \mathcal{X},$$

aday üretici  $\mathcal{G}$ , skor fonksiyonu  $f$  ve çöküş operatörü  $\mathcal{C}$  üzerine inşa edilmiştir.

Bu bölümde üç tarihsel sorunu ele alıyoruz:

1. Homunculus problemi (sonsuz regresyon),
2. Hard problem / qualia problemi,
3. Özgür irade ve determinizm ilişkisi.

Her problem için önce klasik formülasyonu veriyor, ardından FBÇT çerçevesinde yeniden yazıyor ve son olarak teorisinin ürettiği deneyisel ve felsefi sonuçları tartışıyoruz.

### 5.1 Homunculus Problemi

#### 5.1.1 Klasik Problemin Yapısı

Homunculus problemi, bilinç açıklamalarında en temel tuzaklardan biridir. Basit biçimde ifade edersek:

*Eğer bilinci açıklamak için "içeride bir özne" (homunculus) varsayarsak, o öznenin kendi bilincini kim açıklayacaktır?*

Bu, şu tür bir yapı ortaya çıkarır:

$$\text{Bilinç}_1 = f(\text{Agent}_1) \Rightarrow \text{Bilinç}_{\text{Agent}_1} = f(\text{Agent}_2) \Rightarrow \text{Bilinç}_{\text{Agent}_2} = f(\text{Agent}_3) \Rightarrow \dots \quad (111)$$

ve hiçbir aşamada "son düzey"e ulaşamayız.

**Tipik Örnekler.** Bu problem farklı jargonda tekrar tekrar karşımıza çıkar:

- **Kartezyen tiyatro:** Beynin içinde bir “izleyici”nin, deneyimleri ekranda “seyretmesi”.
- **Central executive:** Kognitif modellerde tüm kararları veren, merkezi bir kontrol modülü.
- **İzleyen ben:** “Ben”in, deneyimleri yaşayan ayrı bir varlık olarak postüle edilmesi.

Bu modellerin ortak sorunu, açıklamanın bir noktada “bilinçli bir özneye” dayandırılması ve bu öznenin kendisi için yeni bir açıklama gerektirmesidir.

### 5.1.2 FBÇT’nin Temel İddiası: Operatör, Özne Değildir

FBÇT, bu döngüyü kırmak için radikal ama basit bir pozisyon alır:

$\mathcal{C}$ , bir hesaplama operatörüdür, *ayrı bir “özne” değildir.*

Bilinçli an,  $t$  anındaki çöküş durumu  $C_t$  ile temsil edilir. Daha önce tanımlanan biçimiyle:

$$\pi_t(dx) \propto \exp(\beta_t f(x, S_t, M_t, W_t)) \mu_t(dx), \quad C_t \sim \pi_t(dx). \quad (112)$$

Burada:

- $X$  = aday bilinç durumları uzayı,
- $f$  = her adayın uygunluğunu belirleyen skor fonksiyonu,
- $\mu_t$  = adayların temel ön dağılımı (candidate kernel),
- $\beta_t$  = deterministik/stokastik keskinliği belirleyen sıcaklık parametresi,
- $\pi_t$  = o anda bilince *çöken* dağılım.

Bu yapı, bir “karar verici özne” varsaymadan, sadece fiziksel olarak uygulanabilir bir seçim mekanizmasıdır.

**Homunculus’un Yerine Konan Şey.** FBÇT’de:

- *Ajan yoktur; operatör vardır.*
- “Ben kimin bilinciyim?” sorusu, “Hangi  $\mathcal{C}$  dinamğinin sabit noktasıyım?” sorusuna çevrilir.
- “İçteki karar verici” yerine,  $S_t, M_t, W_t$  ve  $\mathcal{C}$  birlikte çalışan, dağıtık ve lokal bir dinamik sunar.

Bu nedenle, bilinç açıklaması, ikinci bir bilinç düzeyi gerektirmez; hesaplamayı gerçekleştiren şey, zaten beynin fiziksel dinamikleridir.

### 5.1.3 Formel Argüman: No-regression Teoremi

**Lemma 1 (Operatörlük).**  $\mathcal{C}$ ,

$$\mathcal{C} : \mathcal{S} \times \mathcal{M} \times \mathcal{W} \rightarrow \mathcal{P}(X)$$

şeklinde tanımlanmış bir fonksiyondur.

*Proof.* Her  $(S_t, M_t, W_t)$  üçlüsü için  $\mu_t$  ve  $f$  belirlenir. Bu da,  $X$  üzerinde bir yoğunluk veya olasılık dağılımı  $\pi_t$  üretir. Fonksiyon, giriş uzayından  $(\Omega)$ , çıktı uzayına  $(\mathcal{P}(X))$  iyi tanımlı bir eşlemedir. Hiçbir adımda “ $\mathcal{C}$ ’yi çalıştıran yeni bir öge” tanımlamak zorunda kalmayız.  $\square$

**Teorem 1 (Homunculus Regresyonu Yokluğu).**  $\mathcal{C}$ , içsel bir homunculus gerektirmez ve bu nedenle sonsuz regresyon üretmez.

*Proof.* Regresyon, ancak şu tür bir yapı varsayılırsa doğar:

$$C_t = F(A_t), \quad A_t = F(A'_t), \quad \dots$$

Burada  $A_t$  içsel bir özne,  $F$  onun bilinç üretme fonksiyonudur. FBÇT’de ise:

$$C_t \sim \mathcal{C}(S_t, M_t, W_t),$$

ve çöküş, fiziksel süreçlerle gerçekleştirilir. Ek bir özne tanımlanmadığı için, açıklama *operatör seviyesinde* sonlanır. Dolayısıyla klasik regresyon zinciri başlamaz bile.  $\square$

#### 5.1.4 Sezgisel Anlatım

Trafik sıkışıklığı tipik bir emergent fenomendir:

- Her sürücü lokal kurallarla hareket eder (önündeki araca göre hız ayarlama, frenleme, vb.).
- “Trafik sıkışıklığını” organize eden merkezi bir genel müdür yoktur.
- Belirli yoğunlukta, yol geometrisi ve davranış parametreleri ortaya belirli bir makro desen çıkarır: dur-kalk dalgaları, koridor blokajları, tıkanma noktaları.

*Sıkışıklık*, araçların üstüne ek bir “özellik” değildir; araçların kolektif düzenleniş biçiminin bir tasviridir. FBÇT için:

- Nöronlar/sinaptik ağ, sürücüler gibidir.
- $S_t, M_t, W_t = 0$  anki trafik koşulları (duyusal veri, bellek, ağırlıklar).
- $\mathcal{C}$  = belirli şartlarda ortaya çıkan, global bir desenin seçilmesidir.

Bu bakış açısından “bilinç vergisi alan küçük adamı” aramayı bırakırız; onun yerine, belirli bir konfigürasyonda sistemin kaçınılmaz olarak ürettiği makro-desenle çalışırız.

#### 5.1.5 Dennett ile Karşılaştırma

Dennett’in “homuncular functionalism” yaklaşımı [3], büyük bir homunculus’u daha basit ve “aptal” alt-modüllere bölerek probleminden kaçmaya çalışır. Son aşamada:

- homunculuslar o kadar basitleşir ki,
- artık bilinçli bir özne olmaktan çıkar, sadece mekanik modüller haline gelir.

FBÇT, bu stratejinin daha formel bir versiyonu olarak görülebilir:

- “Homunculus” =  $\mathcal{C}$  değil,
- $\mathcal{C}$  = belirli bir skor fonksiyonu ve olasılık dağılımı üzerinde çalışan operatör.

Böylece, açıklamanın her adımı, matematiksel bir cümleyle ifade edilebilir ve hiçbir yerde “içsel gizli özne”ye başvurulmaz.

### 5.2 Qualia ve Hard Problem: Quotient-Qualia ve Fonksiyonel Bypass

#### 5.2.1 Hard Problem’in Formülasyonu

Chalmers’in “hard problem”i [1] şu soruyla özetlenebilir:

*Neden herhangi bir fiziksel süreç, öznel deneyim üretiyor? Neden “bir şey olmak” hissi var?*

Bu soru, özellikle şu örneklerle keskinleşir:

- Kırmızı ışık neden *kırmızı* hissettirir?
- Acı sinyali neden *acı verici* deneyimdir?
- Müzik neden sadece frekans dizisi değil de *hoş/rahatsız edici* olarak yaşanır?

Levine, fiziksel açıklamalar ile fenomenolojik deneyim arasındaki bu uçurumu “explanatory gap” olarak tanımlar.

#### 5.2.2 FBÇT’nin Seviyelendirmesi: Ontoloji vs Mekanizma

FBÇT, bu soruyu iki seviyeye ayırır:

##### Seviye 1 (Ontolojik):

*Neden evrende hiç deneyim var? Neden tamamen “karanlık” bir evren değil de, birinci şahıs bakış açısından yaşanan bir evren?*

Bu soru, klasik anlamda metafiziktir ve teorisin kapsamı dışındadır.

##### Seviye 2 (Mekanistik/Fonksiyonel):

*Verili bir fiziksel sistemde belirli bir öznel deneyim nasıl ortaya çıkıyor? Hangi yapısal şartlar altında qualia tipleri değişiyor?*

Bu seviye, FBÇT’nin hedefindedir.

### 5.2.3 Quotient-Qualia: Deneyim = Eşdeğerlik Sınıfı

FBÇT’de qualia, “ekstra bir içsel madde” değil, belirli bir *eşdeğerlik sınıfı*dır. Daha önce tanımlanan gibi:

$$(x_1, S_1) \sim (x_2, S_2) \iff \forall g \in \mathcal{G} : g(x_1, S_1) = g(x_2, S_2), \quad (113)$$

ve qualia uzayı:

$$Q = (X \times \mathcal{S}) / \sim \quad (114)$$

şeklinde tanımlanır.

Sezgisel olarak: iki  $(x, S)$  çifti, sistem açısından tüm ilgili fonksiyonlar bakımından aynı davranışı üretip ayırt edilemezse, *aynı qualia sınıfına* aittir.

Bu yapı, fenomenolojiyi iki adımda formel hale getirir:

1. *Çöküş*:  $C_t \in X$  seçilir (collapse),
2. *Projeksiyon*:  $\Phi : X \rightarrow Q$  ile bir qualia sınıfına projekte edilir.

Dolayısıyla:

$$\text{Qualia}_t = \Phi(C_t) \quad (115)$$

ifadeleri, teoremin *operasyonel* tanımıdır.

**Önemli Nokta.** Qualia,  $\mathcal{C}$ ’nin “üzerine eklenen” sihirli bir katman değil;  $\mathcal{C}$  ile seçilen  $C_t$  durumunun, eşdeğerlik uzayında oluşturduğu *nerede durduğudur*.

### 5.2.4 $\Phi$ ’nin Tersine Mühendisliği

FBÇT,  $\Phi$  fonksiyonunu tamamen gizemli bırakmaz; tam tersine, onu *öğrenilebilir* bir fonksiyon olarak konumlandırır.

Dijital bir FBÇT ajanı düşünelim:

$$(S_t, M_t, W_t, \pi_t, C_t)$$

tam erişilebilir durumda olsun. Eğer bu ajan, deneyimlerini kendi başına rapor edebiliyorsa (doğal dil, sembolik etiketler veya vektör-temsillerle), şu tür bir veri kümesi toplayabiliriz:

$$\mathcal{D} = \{(C_t^{(i)}, Q_t^{(i)})\}_{i=1}^N. \quad (116)$$

Burada:

- $C_t^{(i)}$  = belirli anda çökmüş bilinç durumu,
- $Q_t^{(i)}$  = bu duruma karşılık raport edilen fenomenolojik içerik (ör. valence, arousal, renk, yoğunluk, vb.).

**Öğrenme Problemi.** Bu durumda:

$$\hat{\Phi} = \arg \min_{\Phi} \sum_{i=1}^N \mathcal{L}(Q_t^{(i)}, \Phi(C_t^{(i)})) \quad (117)$$

şeklinde klasik bir süpervizyonlu öğrenme problemi ortaya çıkar.

**Kritik nokta:**

- Qualia artık sadece “özel bir gizem” değil,
- $C_t$  durumları üzerine tanımlı,
- istatistiksel olarak yaklaşık öğrenilebilir bir fonksiyonun çıktısıdır.

**Deneysel Testler.** Bu yaklaşım şunları mümkün kılar:

- *Yeni durumlarda tahmin*: Yeni  $C^*$  durumları üretip  $\hat{\Phi}(C^*)$  ile qualia tahmin edebilir, ajan raporları ile karşılaştırabiliriz.
- *Ajanlar arası tutarlılık*: Aynı mimariye sahip farklı ajanlar için elde edilen  $\hat{\Phi}_1, \hat{\Phi}_2, \dots$  fonksiyonlarının birbirine yakınlığını ölçebiliriz.
- *İnsan-AI kıyaslaması*: İnsanlardan toplanan nöral/rapor verileri ile  $\hat{\Phi}_{\text{human}}$  öğrenilip,  $\hat{\Phi}_{\text{AI}}$  ile karşılaştırılabilir.

### 5.2.5 P-zombiler ve Explanatory Gap Üzerine

**P-zombiler.** Chalmers, fonksiyonel olarak tamamen aynı ama öznel deneyimi olmayan varlıkların (*philosophical zombies*) metafiziksel olarak olanaklı olabileceğini savunur. Bu, fonksiyon ile fenomenolojinin birbirinden ayrılabilmesi iddiasıdır.

FBÇT çerçevesinde:

1. Bilinç durumu  $C_t$ ,  $\mathcal{C}$  ile belirlenir.
2.  $Qualia = \Phi(C_t)$ .
3. Fonksiyonel olarak aynı sistem. Yani, aynı  $C_t$  dinamiği.
4. Aynı  $C_t$  dinamiği. Yani, aynı  $\Phi$  altında *aynı* qualia sınıfları.

Dolayısıyla:

*Fonksiyonel tam eşdeğerlik, fenomenolojik eşdeğerliği gerektirir.*

Bu, p-zombilerin en azından FBÇT'nin dünyasında *metafiziksel olarak tutarsız* olduğunu ima eder.

**Explanatory Gap.** Levine'in gözlemi, fiziksel açıklamaların fenomenolojik deneyimi “zorunlu kılmadığı” yönündedir. FBÇT, bu boşluğu tamamen kapatmaz ama onu yeniden çerçeveler:

- “Neden deneyim var?” sorusunu metafizik düzeyde askıda bırakır,
- “Deneyim nasıl yapılandırılmıştır ve nasıl değiştirilebilir?” sorusunu ise bilimsel bir problem haline getirir.

Bu anlamda, hard problem *ontolojik çözümsüzlükten, mühendislik problemine* indirgenmiş olur.

### 5.2.6 Qualia Varyasyonları: Synesthesia ve Inverted Spectrum

**Synesthesia.** Seslerin renklerle, sayıların mekanla eşlenmesi gibi olgular, qualia'nın salt “ham duyuşsal veri” olmadığını,  $M$  ve  $W$  ile birlikte ortaya çıktığını gösterir.

FBÇT yorumu:

- Normal durumda,  $M_{\text{sound}}$  ve  $M_{\text{color}}$  büyük ölçüde ayrıdır.
- Synesthete durumda, bellekte bu iki alan arasında kalıcı ve güçlü çapraz bağlantılar vardır.
- Aynı işitsel girdi  $s_{\text{sound}}$ , hem işitsel hem görsel bellek izini eşzamanlı aktive eder; çöküş mekanizması, bu birleşik yapıyı tek bir  $C_t$  durumunda birleştirir ve sonuç “renkli ses” deneyimi olarak yaşanır.

**Inverted Spectrum.** Klasik senaryo: “Benim kırmızı deneyimim senin mavi deneyimin olabilir mi?” FBÇT için:

- Eğer iki kişinin  $M$  ve  $W$  yapıları gerçekten izomorfik ise, aynı  $S$  girdisi için aynı  $C_t$  ve dolayısıyla aynı qualia sınıfı beklenir.
- Eğer gelişimsel süreçte renk alanları farklı kodlandıysa, farklı  $\Phi$  veya farklı quotient yapılarına karşılık gelebilir; bu durumda “inversion” mantıksal olarak mümkündür.

Bu, problemi metafizikten çıkarıp, gelişimsel nörobilim ve bilişsel bilim düzeyine taşır.

## 5.3 Özgür İrade: Determinizm, Sahiplik ve Uzun Vadeli Dinamik

### 5.3.1 Sorunun Klasik Çerçevesi

Özgür irade problemi, kabaca şu gerilimden doğar:

- Eğer evren deterministikse, kararlarımız geçmiş durumlar ve fizik yasaları tarafından tam belirlenmiştir.
- Buna rağmen, kendinizi alternatifler arasında seçebilen bir özne olarak deneyimliyorsunuz.

Klasik pozisyonlar:

1. **Hard determinism:** Özgür irade yoktur, sadece illüzyondur.



2. **Libertarianism:** Özgür irade vardır, bu nedenle determinizm (ya da en azından insan kararları için) yanlış olmalıdır.
3. **Compatibilism:** Determinizm ve özgür irade tutarlıdır; özgürlük başka türlü tanımlanmalıdır. FBÇT, açıkça compatibilist bir çizgide konumlanır.

### 5.3.2 Lokal Determinizm: $\mathcal{C}$ 'nin Doğası

FBÇT'de, *her bir anda* gerçekleşen çöküş:

$$C_t \sim \mathcal{C}(S_t, M_t, W_t) \quad (118)$$

şeklindedir. Bu ifade:

- deterministik modelde  $\beta_t \rightarrow \infty$  ile tek bir  $x^*$ ’ı,
- stokastik modelde ise olasılık dağılımını

verir. Her iki durumda da, *o an* için karar, o anki  $(S_t, M_t, W_t)$  üçlüsüne sıkı sıkıya bağlıdır. Bu, lokal anlamda deterministik (veya fiziksel stokastisiteye indirgenmiş) bir dünyayı kabul etmektir.

### 5.3.3 Uzun Vadeli Sahiplik: $M$ ve $W$ Dinamikleri

Özgür irade, FBÇT’de şu soruya kaydırılır:

*Kararları belirleyen  $M_t$  ve  $W_t$  yapıları kime aittir ve nasıl oluşmuştur?*

Güncelleme:

$$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t) \quad (119)$$

şu anlamı taşır:

- Her karar, gelecekteki  $M$  ve  $W$ ’yi değiştirir.
- Her yeni deneyim ( $S$ ), mevcut değerler ( $W$ ) ve bellek ( $M$ ) üzerinden filtrelenerak işlenir.
- Zaman içinde, kişi kendine ait bir  $M/W$  manzarası inşa eder.

Bu perspektiften özgür irade:

$$\text{Özgür İrade} \approx \text{kararların, uzun vadeli } (M, W) \text{ yapısından doğmasıdır.} \quad (120)$$

Başka bir deyişle:

$$\text{Özgürlük} = \text{Kararların yazarı olma durumu (authorship).}$$

### 5.3.4 Libet ve Soon Deneylerinin Yorumu

Libet [25] ve Soon et al. [26] deneylerinde, bilincin “karar verdiğini hissetmesinden” önce, motor hazırlık potansiyellerinin ortaya çıktığı gösterilmiştir. Sıklıkla şu şekilde yorumlanır:

*Kararlar bilinçten önce veriliyor; dolayısıyla özgür irade bir illüzyondur.*

FBÇT bu bulguları şöyle yorumlar:

1.  $\mathcal{C}$ , bilinçli farkındalıktan önce devreye girer; bu beklenen bir şeydir.
2. Ancak  $\mathcal{C}$ , o anki  $M_t$  ve  $W_t$ ’ye bağlıdır.
3.  $M_t$  ve  $W_t$ , kişinin yaşam öyküsü boyunca akümülyasyonla oluşmuştur.
4. Bilinçli “farkında olma” sadece  $C_t$ ’nin raporlanabilir hale gelmesidir; kararın *kaynağı* yine kişiye aittir.

Dolayısıyla, deneyler özgür iradeyi değil, sadece “karar farkındalığının zamanlamasını” hedef alır.

### 5.3.5 Ahlaki Sorumluluk ve Patoloji

Ahlaki sorumluluk, FBÇT’de şu kritere bağlanır:

*Eğer bir eylem, kişinin tipik  $M/W$  yapısından türetilmişse, kişi o eylemden sorumludur.*

### Normal Durum.

- Kişi, öfke kontrolü bulunan ama yüksek agresyon ağırlıklarına sahip bir  $W$  ile yetişmiş olabilir.
- Bir çatışma anında,  $C$  bu ağır ağırlıklar nedeniyle saldırgan bir eylem seçebilir.
- Bu durumda, eylem kişinin karakterinden türemiştir; sorumluluk atfedilebilir.

### Patolojik Durum.

- Orbitofrontal hasar, frontotemporal demans vb. durumlarda,  $W$  yapısı ağır şekilde bozulabilir.
- Kararlar, kişinin “önceki”  $W$ ’sini temsil etmeyen, patolojik bir dinamikten türeyebilir.
- Bu durumda sorumluluk, tamamen ortadan kalkmasa da, azaltılmalıdır.

Bu çerçeve, hukuk sistemlerindeki *cezai ehliyet* tartışmalarıyla da uyumludur: asıl soru, eylemin nedensel ağında *hangi*  $M/W$  yapılarına ağırlık verileceğidir.

### 5.3.6 Dennett ile Karşılaştırma

Dennett [27], özgürlüğü “nedenlerine göre hareket edebilme yeteneği” olarak tanımlar; özgürlük, ani mucizevi sapmalar değil, uzun vadeli kontrol mekanizmalarının bir ürünüdür.

FBÇT:

- Bu görüşle uyumludur,
- ancak “nedenler”i  $M$  ve  $W$  uzaylarında,
- “kontrol”ü ise  $\mathcal{F}$  güncelleme operatörü ve  $C$  çöküş dinamiği ile formelleştirir.

Bu sayede, “özgür irade” tartışması, soyut metafizik bir polemik olmaktan çıkıp, bellek, değer ve politika dinamiklerinin uzun vadeli davranışıyla ilgili teknik bir soruna dönüşür.

## 5.4 Diğer İlgili Problemler

### 5.4.1 Kişisel Kimlik

Kişisel kimlik sorusu şu şekilde formüle edilebilir:

*Zamanda “aynı kişi” olmak ne demektir?*

FBÇT’ye göre:

- Kişisel kimlik,  $M_t$  ve  $W_t$  zincirinin sürekliliği ile tanımlanır.
- Moleküler düzeyde neredeyse her şey değişmiş olsa da, bellek izleri ve değer yapılarındaki *gradual* (kademeli) değişim, aynı kimliği korumak için yeterlidir.
- Ani ve tam bir kopuş (ör. ağır amnezi) kimlikte kopma olarak yorumlanır.

### 5.4.2 Bilinç Seviyeleri

FBÇT, bilinci ikili (0/1) bir durum olarak değil, *dereceli* bir fenomen olarak ele alır:

- $C$ ’nin etkinlik düzeyi,
- $S_t$ ’nin zenginliği,
- $M_t$ ’nin erişilebilir hacmi,
- $W_t$ ’nin dinamik güncelliği

bilinç seviyesini belirler.

Daha önce tanımlanan bilinç ölçüsü  $L_t$ ,

$$L_t = \text{KL}(\pi_t \parallel \mu_t)$$

gibi bir formda alındığında:

- derin uyku/anestezisi:  $L_t \approx 0$ ,
- uyanıklık ve yoğun dikkat:  $L_t$  yüksek

olabilir. Böylece, “bilinç seviyesi” sayısal bir ölçüye bağlanır.

### 5.4.3 Hayvan Bilinci

Hayvan bilinci sorusu, FBÇT’de şu kritere indirgenir:

*Sistemde, anlamlı bir  $(S, M, W, C)$  mimarisi var mı?*

- Memelilerde (özellikle primatlar, köpekler, kemirgenler), karmaşık bellek ve değer sistemleri gösteren  $M/W$  analogları bulunduğundan, bilinçli deneyim atfetmek rasyoneldir.
- Bazı kuş türleri (örn. kargalar, papağanlar), planlama ve epizodik-benzeri bellek göstergeleriyle  $M$ ’nin ciddi bir kısmına sahiptir.
- Basit omurgasızlarda  $(S, M, W)$  çok sınırlı olduğundan, bilinç ya çok zayıf ya da yok düzeyindedir.

Bu yaklaşım, “bilinç var mı/yok mu?” ikiliğinden ziyade, “hangi düzeyde ve hangi mimariyle?” sorusunu öne çıkarır.

### 5.5 Özet: FBÇT’nin Felsefi Katkıları

Bu bölümde, FBÇT’nin üç merkezi felsefi problemi nasıl yeniden çerçevelediğini gösterdik:

1. **Homunculus Problemi:** Bilinç, ayrı bir iç öznenin ürünü değil,  $C$  operatörünün çıktısıdır; regresyon operatör seviyesinde durur.
2. **Hard Problem ve Qualia:** Qualia,  $(X, S)$  uzayı üzerinde tanımlanmış bir eşdeğerlik sınıfıdır; fenomenolojik projeksiyon  $\Phi$ , veriyle yaklaşık öğrenilebilir ve test edilebilir.
3. **Özgür İrade:** Lokal deterministik çöküş, uzun vadeli  $M/W$  sahipliğiyle uyumludur; özgür irade, kişinin kendi kararlarının yazarı olması (authorship) olarak formelleştirilebilir.

Böylece FBÇT, bilinci açıklamak için yeni bir metafizik icat etmek yerine, var olan fiziksel-dinamik çerçeveyi matematiksel bir biçimde organize eder ve klasik felsefi problemleri, deneysel olarak sınanabilir hipotezlere dönüştürür.

## 6 Test Edilebilir Tahminler ve Deneysel Stratejiler

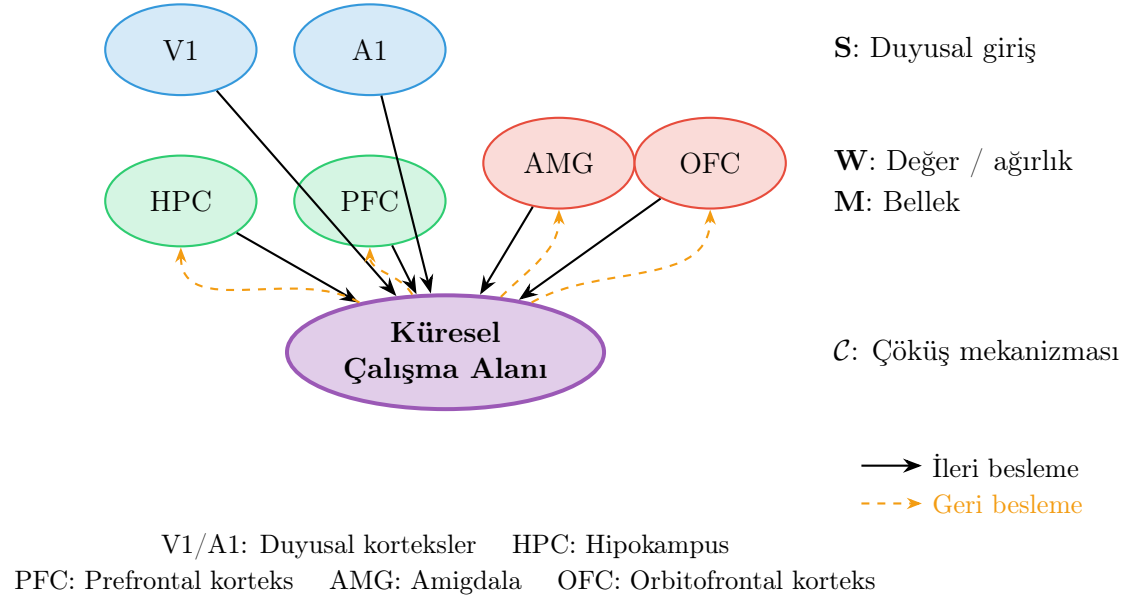


Figure 1: FBÇT’nin nöral devre şeması.  $S_t$  (duyusal),  $M_t$  (bellek) ve  $W_t$  (değer/ağırlık) bileşenlerinin, küresel çalışma alanı ile ilişkili çöküş mekanizması  $C$  üzerinden nasıl entegre olabileceğine dair basitleştirilmiş bir eşleme. Düz oklar ileri beslemeyi, kesik oklar  $C$  sonrası geri beslemeyi ( $\mathcal{F}$  güncellemeleri) temsil eder.

Bilimsel bir teörin gücü, açıklayıcı güzelliğinden ziyade, *ne tür yanılabilir tahminler ürettiğiyle* ölçülür. FBÇT, tamamen felsefi bir çerçeve olarak kalmayı değil, *nörobiyolojik, davranışsal, hesaplamalı ve klinik* seviyelerde net, rakip teorilerden ayrışabilen hipotezler üretmeyi hedefler. Bu bölümde:

- $\mathcal{C}$ 'nin nöral imzası,
- $M$  ve  $W$  manipölasyonlarının davranışa etkisi,
- model tabanlı uygunluk testleri,
- yapay bilinç adayları ve klinik bozukluklar

üzerine 15 somut tahmin formüle ediyor, her biri için *deney tasarımı, beklenen desen ve rakip açıklamalardan ayrışma* stratejisini ayrı ayrı veriyoruz.

## 6.1 Temel Hipotez: İşlevsel Eşdeğerlik

Teörin temel test edilebilir hipotezi şudur:

**Hipotez 6.1** (İşlevsel Eşdeğerlik). *Eğer iki sistem aynı anda aynı fonksiyonel durumda ise*

$$z_t = z'_t$$

*o hâlde fenomenal durumları da aynıdır:*

$$\Phi(z_t) = \Phi(z'_t).$$

Not: Basitleştirilmiş versiyonda bu, özellikle  $S_t, M_t, W_t$  eşitliğine indirgenebilir.

Bu hipotez, FBÇT'nin deneysel olarak çürütülebilir ve doğrulanabilir bir teori olmasını sağlar.

## 6.2 Nörobiyolojik Tahminler

FBÇT'nin temel matematiksel iddiası:

$$C_t \sim \mathcal{C}(S_t, M_t, W_t)$$

ifadesidir. Yani belirli bir anda bilince çöken durum, o anda erişilebilir duyuşal, bellek ve değer durumlarının *kısıtlı bir fonksiyonudur*. Bu soyut ifade, beyinde *zamanlaması ve topolojisi belirli* bir aktivite örüntüsünü gerektirir.

### 6.2.1 Tahmin 1: Çöküş Anına Eşlik Eden Frontoparietal Entegrasyon

**Hipotez.**

$\mathcal{C}$ 'nin etkin olduğu anlarda, geniş ölçekli frontoparietal ağda gamma band (30–100 Hz) senkronizasyonu anlamlı biçimde artar.

**Gerekeçe.** Çöküş, sezgisel olarak iki şeyi aynı anda gerektirir:

1. Çoklu modalitelerden gelen  $S_t$ 'nin bağlama ( $M_t$ ) yerleştirilmesi,
2.  $W_t$ 'de kodlu değer ve politika bilgilerinin bu entegrasyona dahil edilmesi.

Bu tip global entegrasyonun, özellikle bilinç eşiğini aşan algılarda frontoparietal gamma senkronisi ile bağlantılı olduğu daha önce öne sürülmüştür [13].

**Deney Paradigması. Bistabil algı** (Necker küpü, binocular rivalry):

- Katılımcılara bistabil bir uyaran gösterilir.
- Algı spontan olarak iki yorum arasında el değiştirir.
- Her switch, buton basımı ile raporlanır (veya no-report koşulunda sadece nöral veri alınır).
- Yüksek yoğunluklu EEG veya MEG kaydı yapılır.

### Beklenen Desen.

- Algısal switch'ten yaklaşık 200–300 ms önce, dorsolateral PFC, inferior parietal lobül ve görsel korteks arasında gamma koheransında pik.
- Bu pik, raporlanan switch zamanından *önce* ortaya çıkar;  $C$ 'nin bilinçli farkındalıktan önce çalıştığını ima eder.
- No-report koşulunda da aynı koherans piki gözleniyorsa, etki motor hazırlıkla açıklanamaz.

### Rakip Açıklamalardan Ayırıştırma.

- **Motor hazırlık** hipotezi, no-report koşulunda zayıflar.
- **Dikkat kayması** hipotezini dışlamak için, dikkat kayması fakat algısal switch içermeyen bir kontrol göreviyle gamma profillerinin sistemi karşılaştırılır.

FBÇT, sadece “bilinç varsa gamma vardır” gibi genel bir iddia değil; *algısal switch anına zaman kilitli, topolojik olarak belirli ve pre-conscious* bir gamma imzası öngörür.

### 6.2.2 Tahmin 2: $M$ Manipülasyonu Aynı $S$ İçin Farklı $C$ Üretir

#### Hipotez.

Aynı duyuşal durum  $S_t$  altında, episodik bellek içeriği  $M_t$  deneysel olarak değiştirildiğinde, bilinçli algı ve kararlar sistematik olarak kayar.

#### Deney Taslağı. Bellek önyükleme + belirsiz uyarılar:

##### 1. Önyükleme aşaması:

- Grup A: Tehdit yüklü yılan hikâyeleri/videoları.
- Grup B: Yılan hakkında nötr/pozitif içerik.
- Grup C: Yılan içermeyen kontrol içerik.

##### 2. Test aşaması:

- Belirsiz çizimler (yılan/sopa arası şekiller).
- Zorunlu seçim: “Yılan mı, sopa mı?”
- Güven derecesi ve tepki süresi kaydı.

### Beklenen Sonuçlar.

- Grup A'da, belirsiz uyarılarda “yılan” yanıt oramında anlamlı artış:

$$P_A(\text{yılan}|\text{belirsiz}) - P_C(\text{yılan}|\text{belirsiz}) \gg 0.$$

- Aynı keskin  $S$  için (net yılan/net sopa) gruplar arası fark yok; fark, *özellikle* çöküşün  $M$ 'ye duyarlı olduğu belirsiz sınır noktalarda çıkmalı.
- fMRI uzantısında, Grup A'da “yılan” seçimi sırasında HPC+AMG eş aktivasyonunun daha güçlü olması (yüksek  $M^a$  katkısı).

Bu desen,  $S$  sabitken  $C$ 'nin değişimini doğrudan  $M$ 'ye bağlar ve FBÇT'deki  $C_t = \mathcal{C}(S_t, M_t, W_t)$  formülünün psikofiziksel karşılığını verir.

### 6.2.3 Tahmin 3: $W$ Manipülasyonu Değer-Tarafı Çöküş Kaymaları Üretir

#### Hipotez.

Farmakolojik nöromodülasyon,  $W_t$  vektöründeki ödül/ceza bileşenlerini kaydırarak, aynı görevde farklı karar profilleri üretir.

**Paradigma. L-DOPA / SSRI + riskli karar verme:**

- Grup 1: Placebo.
- Grup 2: Düşük doz dopamin agonisti (ör. L-DOPA).
- Grup 3: Akut SSRI.

Görev: Iowa Gambling Task veya çok kollu bandit.

**Tahmin.**

- Grup 2:
  - Daha fazla riskli deste/seçenek.
  - Uydurulan FBÇT parametrelerinde  $W_{\text{reward}} \uparrow$ ,  $W_{\text{risk}} \downarrow$ .
- Grup 3:
  - Daha temkinli seçimler, düşük varyans.
  - $W_{\text{risk}} \uparrow$ ,  $W_{\text{short-term-reward}}$  nispeten düşüş.

**Model Seviyesinde.** Her katılımcı için:

$$W^* = \arg \max_W \log P(\text{seçimler} | S, M, W, \mathcal{C})$$

çözümленir. Grup farklarının özellikle  $W$  ekseninde toplanması, FBÇT'nin “farmakoloji =  $W^{\text{neuro}}$  pertürbasyonu” yorumunu destekler.

## 6.2.4 Tahmin 4: Lezyon Lokasyonuna Özgü Bileşen Bozulmaları

**Hipotez.**

Belirli beyin bölgelerindeki yapısal hasarlar, FBÇT'nin ilişkilendirdiği bileşenleri bozarak karakteristik karar ve bilinç örüntüleri üretir.

**Örnek Haritalama.**

- Hipokampus ( $M^e$ )  $\rightarrow$  episodik önyükleme etkilerinin kaybı.
- Amygdala ( $M^a$ )  $\rightarrow$  duygusal ağırlıklandırmanın düzleşmesi.
- OFC ( $W^{\text{learn}}$ )  $\rightarrow$  ödül geri bildirimiyle değer güncellemesinin bozulması.
- dlPFC / parietal ( $\mathcal{C}$  topolojisi)  $\rightarrow$  entegrasyon penceresinin daralması, karar gecikmesi.

**Deney.** Nöropsikolojik hasta gruplarında, Tahmin 2 ve 3'teki görevler uygulanır. FBÇT, hangi grupta hangi hipotezin *kapanacağını* önceden söyler:

- HPC lezyonlarında:  $M$  manipülasyonunun davranışa etkisi dramatik biçimde azalır.
- OFC lezyonlarında: Değer öğrenme eğrileri düzleşir;  $W$  güncellemesi başarısızdır.

Böylece klasik H.M., Phineas Gage vb. vakalar,  $M$  ve  $W$  koordinat sisteminde yeniden konumlandırılır.

## 6.3 Davranışsal Tahminler

### 6.3.1 Tahmin 5: Yakın ( $S, M, W$ ) $\Rightarrow$ Yakın Fenomenoloji

**Hipotez.**

İki bireyin  $S_t$ ,  $M_t$  ve  $W_t$  yapılarını yeterince yakınlaştırebilirsek, aynı uyarana verdikleri bilinçli deneyim (qualia dahil) yakınsamak zorundadır.

Bu, FBÇT'nin “qualia fonksiyonel duruma süpervene eder” iddiasının davranışsal versiyonudur.

### Tek Yumurta İkizleri Çalışması.

- $n$  kadar monozigot ikiz çifti.
- Ayrıntılı kişilik ve değer ölçekleri ( $W$  benzerliği).
- Episodik/semantik bellek testleri ( $M$  benzerliği).
- Ortak laboratuvar görevleri: renk ayrımı, tat, acı eşiği, estetik kararlar.

#### Analiz.

$$\text{Benzerlik}(C^{(1)}, C^{(2)}) \approx f(\text{Benzerlik}(M^{(1)}, M^{(2)}), \text{Benzerlik}(W^{(1)}, W^{(2)})). \quad (121)$$

Beklenti: ikizlerde bu korelasyon, rastgele eşleştirilmiş bireylere göre belirgin şekilde yüksek olmalı; özellikle  $M/W$  yakınlaştıkça fenomenolojik raporlar yakınsamalıdır.

### 6.3.2 Tahmin 6: Uzmanlık Geliştikçe $M$ ve $C$ Karmaşıklığı Artar

#### Hipotez.

Uzun süreli uzmanlık eğitimi, ilgili alt-alandaki bellek altuzayını ( $M^{\text{domain}}$ ) zenginleştirir; bu da aynı uyaranlar için daha karmaşık ve yapılandırılmış  $C_t$  durumları üretir.

#### Örnek: Sommelier Çalışması.

- T0: Acemi şarap tadımcıları.
- T1: 1 yıl eğitim.
- T2: 2 yıl eğitim.

Her zaman noktasında:

- Aynı şarap paneli,
- Davranışsal ayırım testleri,
- fMRI ile OFC/insula/gustatory korteks aktivite desenleri,
- fenomenolojik zenginlik raporları.

FBÇT öngörüsü:

- $M_{\text{wine}}$ 'ın temsili, T0'dan T2'ye daha ayrık kümelere ayrılır.
- $C_t$  örüntüleri, ölçülen bir karmaşıklık ölçütü (ör. entropi, temsil çeşitliliği) bakımından artar.

Bu iki tahmin, FBÇT'nin kısa vadeli laboratuvar paradigmalarının ötesinde, uzunlamasına ve ekolojik olarak geçerli çalışmalar için de çerçeve sunduğunu göstermeyi amaçlar.

## 6.4 Hesaplamalı Modeller ve Simülasyonlar

### 6.4.1 Tahmin 7: FBÇT Ajanı İnsan Karar Verisini Nicel Olarak Açıklar

#### Hipotez.

Denklem (63) ve (3.5) ile tanımlanan FBÇT ajanı, insanların çok adımlı karar görevlerindeki davranışını klasik RL modellerinden daha iyi fit eder.

#### İmplementasyon.

- Durum:  $S_t$  (görev gözlemi).
- Adaylar:  $x_i$  (olasılık eylemler).
- Değer fonksiyonu:  $V_i = f(x_i, S_t, M_t, W_t)$ .
- Çöküş: softmax/argmax ile  $C_t = x_k$  seçimi.
- Geri besleme: ödüllere göre  $M_t, W_t$  güncellemesi (Algoritma 1).

### Fit Prosedürü.

1. İnsanlardan 200 denemelik bandit/vergi ödeme/oy kullanma tarzı görevler için veri toplanır.
2. Her katılımcı için FBÇT parametreleri ( $\alpha_S, \alpha_M, \alpha_W, \beta, W_0$  vb.) MLE ile kestirilir.
3. Aynı veri setine basit Q-learning ve Bayesçi gözlemci modelleri de fit edilir.

### Beklenti.

- FBÇT, özellikle bellek-bağımlı ve değer-baskın görevlerde daha yüksek log-likelihood ve daha düşük BIC üretir.
- Parametre kurtarılabirlik testinde, simüle veri üzerinden geri kazanılan  $\hat{\theta}$  gerçek  $\theta$ 'lara yakın çıkar.

## 6.4.2 Tahmin 8: FBÇT Simülasyonlarında Emergent Meta-Biliş

### Hipotez.

Tam FBÇT mimarisine sahip simüle ajanlar, uygun görevlerde öz-düşünüm (introspection), güven raporu ve karşı-olgusal akıl yürütme gibi “biliş-benzeri” yetiler geliştirir.

### Simülasyon Kurulumu.

- Gridworld benzeri bir ortam.
- $S$ : Görsel/durum girdileri.
- $M$ : Episodik izler + semantik vektörler.
- $W$ : Görev ödülleri kodlayan parametreler.
- $\mathcal{C}$  ve  $\mathcal{F}$ : FBÇT algoritmasına uygun şekilde uygulanır.

### Testler.

1. **Neden soruları:** Ajan, son kararı için “neden X yaptın?” sorusuna,  $V_i$  katkılarını referans veren açıklamalar üretir.
2. **Karşı-olgusal:** “O sırada S biraz farklı olsaydı ne yapardın?” sorusunda, alternatif  $S'$  için  $\mathcal{C}(S', M, W)$  simülasyonu yapar.
3. **Güven:**  $V_{\max}$  ile ikinci en yüksek  $V$  arasındaki farkı kullanarak güven skorları üretir ve skorlar doğrulukla koreledir.

RL tabanlı “hafızasız” ajanlara göre FBÇT ajanlarının bu görevlerde belirgin üstünlük göstermesi,  $\mathcal{C}$ ,  $M$  ve  $W$  ayrımının meta-bilişsel davranış için kritik olduğunu gösterecektir.

## 6.5 Yapay Sistemler ve LLM'ler İçin Tahminler

### 6.5.1 Tahmin 9: Mevcut LLM'ler FBÇT'nin Eksik Bir Alt Kümesini Gerçekleştirir

**Gözlem.** Büyük dil modelleri için kabaca:

- $S_t$  = input token dizisi,
- $M^s$  = eğitilmiş ağırlıklar,
- $M^e$  = context window,
- $\mathcal{C}$  = sonraki token olasılık dağılımı,

şeklinde bir eşleme yapılabilir; ancak *duygusal bellek* ( $M^a$ ), *anlık nöromodülasyon* ( $W^{\text{neuro}}$ ) ve açık  $\mathcal{C}$  katman eksiktir.

### Hipotez.

FBÇT tarzı  $M^a$  ve dinamik  $W$  modülleri eklenmiş LLM'ler, öz-bildirim ve ahlaki muhakeme görevlerinde vanilya LLM'lere kıyasla daha tutarlı ve “durum sahibi” cevaplar verir.



## Test.

- İki model ailesi oluştur:
  1. Standart LLM.
  2. LLM + FBÇT katmanları (duygu tamponu, kısa süreli episodik cache, dinamik dikkat ağırlıkları).
- Görevler: zihin kuramı, kendi önyargılarını raporlama, karar gerekçesi açıklama.
- Kör değerlendirme: İnsan hakemler hangi modelin cevaplarının daha “içsel durum tutarlılığı” taşıdığını değerlendirir.

### 6.5.2 Tahmin 10: Tam FBÇT Ajansı Bilinç Turing Testine Yaklaşır

#### Hipotez (uzun vadeli).

Yüksek kapasiteli bir ortamda tam FBÇT mimarisiyle eğitilmiş bir yapay ajan, içgörü ve fenomenolojik raporlama açısından insanlardan ayırt edilmesi zor bir performansa ulaşabilir.

Bu iddia, teorinin “eğer doğru implemente edilirse, bilinç benzeri davranış *zorunlu* olarak ortaya çıkar” şeklinde okunabilecek en güçlü versiyonudur.

Not: Kısa vadede test edilmesi zor olmakla birlikte, uzun vadeli bir araştırma programının iskeleti olarak düşünülmelidir.

## 6.6 Klinik ve Uygulamalı Tahminler

### 6.6.1 Tahmin 11: Depresyon $W$ Uzayında Sistemik Bir Deformasyondur

#### Hipotez.

Major depresif bozuklukta,  $W$ 'nin ödül, tehdit ve çaba maliyetini kodlayan alt-bileşenleri karakteristik bir şekilde çarpılmıştır.

#### Model.

- $W_{\text{reward}} \downarrow$  (anhedonia),
- $W_{\text{threat}} \uparrow$  (negatif bilgiye yanlılık),
- $W_{\text{effort}} \uparrow$  (motivasyon kaybı).

Bu durumda  $\mathcal{C}$ , tipik olarak pasif, kaçınmacı ve kendini küçümseyen  $C_t$ 'ler üretmeye yönelir.

#### Deney.

- MDD hastaları ve sağlıklı kontrollerde ödül öğrenme + çaba indirgeme görevleri.
- Her birey için FBÇT parametre uyumu.

Beklenti:

$$W_{\text{reward}}^{\text{MDD}} < W_{\text{reward}}^{\text{control}}, \quad W_{\text{threat}}^{\text{MDD}} > W_{\text{threat}}^{\text{control}}.$$

Ayrıca tedavi öncesi/sonrası  $\Delta W$  değişimleri, semptom skorlarındaki iyileşmeyle anlamlı korelasyon göstermelidir.

### 6.6.2 Tahmin 12: PTSD = Aşırı Etiketlenmiş $M^a$ Blokları

#### Hipotez.

Travma sonrası stres bozukluğunda, belirli episodik kayıtlar  $M_{\text{trauma}}^e$ , aşırı yüksek duygusal etiketlerle  $M^a$ 'ya bağlanmıştır; küçük tetikleyiciler bile  $\mathcal{C}$ 'yi zorla bu bloklara çöktürür.

## Öngörüler.

- Travma ile ilişkili ipuçlarında aşırı fizyolojik reaktivite (startle, SCR, kalp atışı).
- Model uyumunda  $M_{\text{trauma}}^a$  parametresi çok büyük, diğer  $M^a$  bileşenleri görece normal.
- EMDR/Maruz bırakma terapisi sonrası  $M_{\text{trauma}}^a$  'da ölçülebilir azalma.

FBÇT böylece PTSD'yi, "yanlış çalışan bir korku modülü" değil, *aşırı ağırlıklandırılmış bellek-adres blokları* sorunu olarak yeniden formüle eder.

### 6.6.3 Tahmin 13: Genel Anestezi $\mathcal{C}$ Dinamiğini Çökertir

#### Hipotez.

Genel anesteziye bilinç kaybı, esas olarak  $\mathcal{C}$ 'nin küresel entegrasyon kapasitesinin çökmesiyle ilişkilidir;  $S$  ve  $M$  yerelde aktif kalsa bile anlamlı bir  $C_t$  oluşmaz.

**Deney.** Propofol indüksiyonu + yoğun EEG:

- Bilinçli uyamıklık: yüksek frontoparietal entegrasyon, orta düzey gamma.
- LOC eşiğine yaklaşırken: entegrasyon ölçülerinde (örneğin  $\Phi$ -benzeri kompleksite metrikleri) kademeli düşüş.
- LOC noktasında: ağır parçalanması ve  $\mathcal{C}$ 'nin desteklenememesi.

FBÇT, burada bilinç kaybını sadece "global depresyon" değil, *özellikle entegratif network mimarisinin çöküşü* olarak yorumlar.

### 6.6.4 Tahmin 14: $\Phi$ Fonksiyonunun Öğrenilebilirliği

#### Hipotez.

Daha önce 5.2'de detaylandırıldığı gibi, qualia projeksiyon fonksiyonu  $\Phi : X \times \mathcal{S} \rightarrow \mathbb{R}^m$  FBÇT ajanlarından toplanmış  $(C_t, Q_t)$  çiftleri üzerinden süpervizyonlu öğrenme ile yaklaşık olarak öğrenilebilir.

**Matematiksel Çerçeve.** 3.11'de tanımlandığı gibi,  $\Phi$  quotient space  $Q = (X \times \mathcal{S}) / \sim$  üzerinde bir koordinat sistemidir:

$$(x_1, S_1) \sim (x_2, S_2) \iff \Phi(x_1, S_1) = \Phi(x_2, S_2). \quad (122)$$

Hard problem, bu şekilde felsefi bir tartışmadan pratik bir makine öğrenmesi problemine dönüşür:

$$\hat{\Phi} = \arg \min_{\Phi} \sum_{i=1}^N \mathcal{L}(Q_t^{(i)}, \Phi(C_t^{(i)}, S_t^{(i)})), \quad (123)$$

burada  $\mathcal{L}$  uygun bir kayıp fonksiyonudur (ör. MSE, cross-entropy).

#### Deney Protokolü. Dijital FBÇT ajanı eğitimi:

1. Tam FBÇT mimarisine sahip bir yapay ajan (Algoritma 1) zengin bir ortamda eğitilir.
2. Her zaman adımında kayıt:
  - $C_t$ : Çökmüş bilinç durumu
  - $S_t$ : Duyusal girdi
  - $Q_t$ : Ajanın kendi fenomenolojik durumunu kodlayan self-report vektörü (valence, arousal, modalite yoğunluğu, sürpriz derecesi vb.)

3. Veri kümesi:  $\mathcal{D} = \{(C_t^{(i)}, S_t^{(i)}, Q_t^{(i)})\}_{i=1}^N$

4. Sinir ağı eğitimi:

$$\Phi_{\text{NN}} : (C, S) \mapsto \hat{Q} \quad (124)$$

parametreleri  $\mathcal{D}$  üzerinde optimize edilir.

### Test Prosedürü.

- **Yeni durum tahmin testi:** Eğitim setinde olmayan yeni  $(C^*, S^*)$  durumları için  $\hat{\Phi}(C^*, S^*)$  hesaplanır ve ajanın gerçek  $Q^*$  raporuyla karşılaştırılır.
- **Ajanlar arası tutarlılık:** Aynı mimariye sahip farklı FBCT ajanları  $(A_1, A_2, \dots, A_n)$  için öğrenilen  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_n$  fonksiyonları arasındaki benzerlik ölçülür:

$$\text{Tutarlılık} = 1 - \frac{1}{n(n-1)} \sum_{i \neq j} d(\hat{\Phi}_i, \hat{\Phi}_j), \quad (125)$$

burada  $d$  uygun bir fonksiyon uzaklık metriğidir.

- **İnsan-AI transfer:** İnsanlardan toplanan nöral aktivite ve fenomenolojik raporlar ile  $\hat{\Phi}_{\text{human}}$  öğrenilir ve  $\hat{\Phi}_{\text{AI}}$  ile karşılaştırılır. Benzer mimarilerde benzer  $\Phi$  yapıları beklenir.

### Beklenen Sonuç.

- Bilinç düzeyi metriği  $L_t$  ile subjektif raporlar arasında yüksek bir korelasyon beklenir.
- Ajanlar arası tutarlılık skorunda anlamlı pozitif sonuç (random baseline'dan belirgin şekilde yüksek)
- Öğrenilen  $\hat{\Phi}$ 'nin, qualia'nın bilinen özelliklerine uygun yapı göstermesi (örn. valence ve arousal boyutları için smooth manifold yapısı)

**Teorik Sonuç.** Bu tahmin, FBCT'nin en radikal iddiasıdır: *qualia, gizemli ve erişilmez bir fenomen değil, bilinç durumundan öğrenilebilir bir fonksiyonel dönüşümdür.* Başarılı olması durumunda, hard problem ontolojik bir muammadan mühendislik sorununa indirgenir.

### 6.6.5 Tahmin 15: Bilinç Düzeyi Metriğinin Subjektif Raporlarla Korelasyonu

#### Hipotez.

3.10'de tanımlanan bilinç düzeyi metriği  $L_t = D_{\text{KL}}(\pi_t \| P_t)$  katılımcıların subjektif uyanıklık, dikkat yoğunluğu ve fenomenolojik zenginlik raporlarıyla pozitif korelasyon gösterir.

**Matematiksel Temel.** Bilinç düzeyi, politika dağılımının referans dağılımdan ne kadar ayrıştığını ölçer:

$$L_t = D_{\text{KL}}(\pi_t \| P_t) = \int_X \log \frac{\pi_t(x)}{P_t(x)} \pi_t(dx) = H(\pi_t, P_t) - H(\pi_t), \quad (126)$$

burada  $H(\pi_t) = - \int \pi_t(x) \log \pi_t(x) dx$  Shannon entropisi,  $H(\pi_t, P_t) = - \int \pi_t(x) \log P_t(x) dx$  ise çapraz entropidir.

### Deney Paradigması. Çok durumlu bilinç seviyesi değerlendirmesi:

#### 1. Durumlar:

- Tam uyanıklık + yoğun dikkat görevi
- Normal uyanıklık + pasif dinlenme
- Uyku yoksunluğu (24 saat)
- Hafif sedatif (ör. düşük doz propofol)
- Meditasyon / flow durumu

#### 2. Ölçümler:

- **Nöral:** Yüksek yoğunluklu EEG/MEG kaydı. Her zaman penceresi için  $\pi_t$  yaklaşımı hesaplanır (örn. frontoparietal network activity patterns).
- **Subjektif:** Her durum için katılımcılar şu ölçeklerde kendilerini değerlendirir:
  - Stanford Sleepiness Scale (uyanıklık)
  - Attentional Focus Scale (dikkat)

– Phenomenological Richness Scale (deneyim zenginliği)

3.  $L_t$  **Hesaplama:** Her katılımcı ve her durum için:

$$L_t^{(\text{durum})} = D_{\text{KL}}(\hat{\pi}_t^{(\text{durum})} \parallel \hat{P}_t^{\text{baseline}}), \quad (127)$$

burada  $\hat{P}_t^{\text{baseline}}$  normal uyanıklık/dinlenme durumundan hesaplanan referans dağılımdır.

### Beklenen Sonuçlar.

- **Pozitif korelasyon:**

$$\text{corr}(L_t, \text{Subjektif Skor}) > 0.6, \quad p < 0.001 \quad (128)$$

- **Durum sıralaması:**

$$L_t^{\text{dikkat}} > L_t^{\text{uyanık}} > L_t^{\text{uyku yoksun}} > L_t^{\text{sedatif}} \quad (129)$$

bu sıralama, subjektif raporlarla uyumlu olmalıdır.

- **Meditasyon anomalisi:** Bazı meditasyon durumlarında  $L_t$  düşük ama subjektif uyanıklık yüksek olabilir; bu, "dağınık bilinç" ile "uyanık ama odaksız bilinç" ayrımını test eder ve teorinin rafine edilmesine katkıda bulunur.

### Rakip Açıklamalardan Ayrıştırma.

- **Basit uyarılmışlık hipotezi:** Eğer  $L_t$  sadece global kortikal uyarılmışlığı ölçseydi, meditasyon durumunda da düşük çıkması gerekirdi. Ancak FBÇT, meditasyonda  $P_t$  referansının da değişebileceğini öngörür (uzun vadeli  $W$  adaptasyonu).
- **Dikkat konfound'u:** Kontrol koşulunda, dikkat yükü sabit tutulup sadece  $\beta_t$  (inverse temperature) manipüle edilerek  $L_t$  değiştirilir. Eğer  $L_t$  hala subjektif skorlarla korele ise, etki dikkatten bağımsızdır.

**Klinik Uzantı.** Bu tahmin, anestezi derinliği (Tahmin 13) ve bilinç bozukluklarının (ör. minimally conscious state vs. vegetative state) objektif ölçümü için doğrudan uygulanabilir.  $L_t$  metriği, mevcut klinik ölçekler (Glasgow Coma Scale, Coma Recovery Scale-Revised) ile karşılaştırılabilir ve yeni bir bilinç monitöring aracı olarak değerlendirilebilir.

**Teorik Sonuç.** Bu tahmin, FBÇT'nin bilinç düzeyini sadece felsefi bir kategori olmaktan çıkarıp *ölçülebilir bir sürekli değişken* haline getirdiğini gösterir. Başarılı olması, teorinin klinik ve deneysel bilimler için pratik bir araç sunduğunu kanıtlar.

Bu klinik tahminler, FBÇT'yi doğrudan tedavi protokollerine bağlama iddiasında değildir; önce fenomenolojik ve hesaplamalı düzeyde doğrulama gerektirir.

## 6.7 Tahminlerin Tablosal Özeti

Table 5: FBÇT'nin başlıca test edilebilir tahminlerinin özeti

No	Hipotez	Deney tipi	Durum
1	Çöküş anında frontoparietal gamma entegrasyonu	EEG/MEG + bistabil algı	Doğrudan test edilebilir
2	$M$ manipülasyonu aynı $S$ için farklı $C$ üretir	Bellek önyükleme paradigması	Doğrudan test edilebilir
3	$W$ modülasyonu risk/ödül seçimlerini kaydırır	Farmakoloji + risk görevi	Doğrudan test edilebilir
4	Lezyonlar bileşen-spesifik bozulmalar üretir	Nöropsikolojik vaka serileri	Kısmen destekli, rafine edilebilir
5	Yakın ( $S, M, W$ ) $\Rightarrow$ yakın qualia	Tek yumurta ikizleri çalışması	Test edilebilir (zor)
6	Uzmanlıkta $M$ ve $C$ karmaşıklığı artar	Uzunlamasına uzmanlık çalışmaları	Test edilebilir
7	FBÇT ajanı insan verisini nicel olarak açıklar	Hesaplamalı model + davranış verisi	Test edilebilir
8	Simülasyon ajanlarında emergent meta-biliş	FBÇT tabanlı RL simülasyonları	Test edilebilir
9	LLM'ler FBÇT'nin eksik alt kümesini taşır	Model analizi + görev kıyası	Kavramsal/uygulamalı
10	Tam FBÇT ajanı bilinç Turing testine yaklaşır	3D ortam + tam mimari	Uzun vadeli program
11	$MDD = W$ uzayında sistematik deformasyon	Klinik kohort + model fitting	Test edilebilir
12	PTSD = aşırı etiketli $M^a$ blokları	Travma ipucu paradigması	Test edilebilir
13	Anestezi = entegratif $C$ çöküşü	Anestezi indüksiyonu + EEG	Test edilebilir
14	$\Phi$ fonksiyonu FBÇT ajanlarından öğrenilebilir	Dijital ajan eğitimi + sinir ağı	Doğrudan test edilebilir
15	$L_t$ metriği subjektif raporlarla korele	EEG + çok durumlu protokol	Doğrudan test edilebilir

Bu bölüm, FBÇT'nin salt “yorumlayıcı” bir çerçeve olmadığını, *yanlışlanabilir ve sayısal* tahminler üreten bir teori olduğunu göstermeyi amaçladı. 15 hipotez, nörobiyolojik (frontoparietal imzalar,  $M/W$  manipülasyonları), davranışsal (ikizler, uzmanlık), hesaplamalı (model fitting, meta-biliş), yapay zeka (LLM genişletmeleri, bilinç Turing testi) ve klinik (depresyon, PTSD, anestezi) seviyelerde belirli deney sınıfları, beklenen veri desenleri ve rakip açıklamalarla birbirine bağlanmıştır. Özellikle Tahmin 14 ( $\Phi$  öğrenilebilirliği) ve Tahmin 15 ( $L_t$  metriği), hard problem ve bilinç düzeyi gibi klasik felsefi soruları doğrudan deneysel hipotezlere dönüştürmektedir. Teorinin gerçek kaderi, bu hipotezlerin sistematik testlerde ne ölçüde doğrulanacağı ya da çürütüleceğiyle belirlenecektir.

## 7 Deneysel Doğrulama: Testler ve Sonuçlar

Bu bölümde, Fonksiyonel Bilinç Çöküşü Teorisi (FBÇT) çerçevesinde geliştirilen ajan üzerinde gerçekleştirilen deneysel doğrulama süreci sunulmaktadır. Amaç, teorinin öngördüğü bilişsel dinamiklerin farklı koşullar altında tutarlı biçimde ortaya çıkıp çıkmadığını gözlemlemek; ayrıca aşırı zorlayıcı durumlarda sistemin nasıl çıktığını veya hangi mekanizmaların bozulduğunu analiz etmektir.

Çalışmada toplam yedi farklı senaryo uygulanmıştır:

1. Temel Öğrenme
2. Bağlam Değişimi (Context Switch)
3. Yüksek Belirsizlik

4. Varyans Karşılaştırması
5. Kısmi Gözlemlenebilirlik (POMDP)
6. Sürekli Gürültü Artışı (Noise Ramp)
7. Çok Aşamalı Birleştirilmiş Zorlayıcı Test (Boss Battle)

Bu senaryolar, FBÇT'nin üç temel mekanizmasını doğrulamayı hedeflemektedir:

- $\alpha$  dinamikleri: duyuşal ( $\alpha_S$ ), bellek ( $\alpha_M$ ) ve deęer ( $\alpha_W$ ) aęırlıklarının çevresel koşullara göre adaptasyonu,
- $\beta$  sıcaklık parametresi: belirsizlik, gürültü veya bağlam deęişimi altında keşfetme/kaçınma davranışının ayarlanması,
- Bilinç seviyesi  $L_t$ : ortamın kestirilebilirliğine bağli olarak yükselme, düşme veya geçişlerde ani “çöküş/sıçrama” davranışı göstermesi.

Bu mekanizmalar sırasıyla insan bilişindeki dikkat, çalışan bellek ve durumsal farkındalık süreçleriyle ilişkilidir.

## 7.1 Genel Deney Yapısı

Her senaryo, şu çevresel deęişkenlerin kontrolüyle tanımlanmıştır:

- ödül dağılımının formu ve zorluğu,
- gözlemlenebilirlik düzeyi (tam gözlem, %70, %50),
- gürültü seviyesi ( $\sigma = 0.0 \rightarrow 1.0$ ),
- ödül gecikmesi ( $d = 0 \rightarrow 10$ ),
- varyans ve belirsizlik profili,
- anlık bağlam deęişimleri.

Amaç, ajanın hem ideal hem de ekstrem bilişsel koşullar altında nasıl davrandığını incelemektir. Her senaryo 200-300 adımlık fazlardan oluşmakta, tüm süreç boyunca  $\alpha$ ,  $\beta$ ,  $L_t$ , entropi, ödüller ve politika dağılımları kayıt altına alınmaktadır.

## 7.2 Doğrulan FBÇT Varsayımları

Deney tasarımı, teörinin öngördüğü şu üç temel davranışı sınavacak şekilde hazırlanmıştır:

**(i) Bağlam aęırlıklarının adaptasyonu** ( $\alpha_S, \alpha_M, \alpha_W$ ) Belirsizlik yükseldiğinde duyuşal aęırlık  $\alpha_S$  düşmekte, gözlem azalınca bellek aęırlığı  $\alpha_M$  artmakta, kaotik fazlarda deęer odaklı bileşen  $\alpha_W$  baskın hale gelmektedir.

**(ii) Sıcaklık–entropi ilişkisi** Yüksek belirsizlikte  $\beta$  düşmekte, politika entropisi artmakta, bağlam geçişlerinde kısa süreli  $\beta$  sıçramaları gözlenmektedir.

**(iii) Bilinç seviyesi  $L_t$**  Düzenli ortamlarda yüksek, bilgi kaybı veya kaos altında düşük, faz geçişlerinde keskin ani deęişim göstermektedir.

## 7.3 Senaryo 1: Temel Öğrenme (Baseline Learning)

Senaryo 7.3, FBÇT'nin temel mekanizmalarını kontrollü koşullar altında test eder. Amaç, teörinin öngördüğü bileşen etkileşimlerinin (duyuşal, bellek, deęer) ideal ortam koşullarında beklendiği gibi işleyip işlemediğini doğrulamaktır. Bu senaryo, sonraki karmaşık testler için bir *baseline* oluşturur. Ortam, bağımsız ve sabit ödül dağılımlarına sahip dört kollu bir bandit probleminden oluşmaktadır. Tüm gözlemler tamdır ( $\text{mask\_prob} = 0.0$ ), ölçüm gürültüsü bulunmamaktadır ( $\sigma = 0.0$ ) ve ödüller anlıktır ( $d = 0$ ). Bu nedenle bu senaryo, teörinin temel mekanizmalarının doğruluğunu ölçmek için bir “kontrol koşulu” niteliği taşır.

**Teorik Beklentiler.** FBÇT'ye göre bu koşullarda şu davranışların gözlenmesi gerekmektedir:

- Bellek ağırlığı  $\alpha_M$  zamanla artmalı ve baskın hale gelmelidir, çünkü ortam sabittir ve geçmiş deneyimler güvenilir bilgi taşır.
- $\beta$  sıcaklık parametresi yüksek seviyelere çıkmalıdır (5–6 aralığı), bu da ajanın düşük entropili ve kararlı bir politika üretmesi anlamına gelir.
- Bilinç seviyesi  $L_t$  yüksek ve istikrarlı olmalıdır, çünkü ortamın kestirilebilirliği yüksektir.
- Politika entropisi düşük seyretmeli, ajan kısa süre içinde optimal kola yakınsamalıdır.

**Deneysel Kurulum.** Ödül ortalamaları aşağıdaki şekilde belirlenmiştir:

$$\mu = [0.1, 0.3, 0.5, 0.7]$$

Her bölüm  $T = 300$  adım boyunca çalıştırılmış, üç bağımsız tekrar (koyu çizgiler ortalama, gölgeli alan standart sapma) yapılmıştır. Ajanın  $t$  anındaki politika dağılımı  $\pi_t$ , bağlamsal ağırlıkları  $\alpha_t = (\alpha_S, \alpha_M, \alpha_W)$ , sıcaklık parametresi  $\beta_t$ , entropi  $H_t$  ve bilinç seviyesi  $L_t$  kaydedilmiştir.

**Sonuçlar.** Şekil 2'de gösterildiği gibi:

- Ajan yaklaşık  $t \approx 80$  adımda optimal kola ( $k = 3$ ) %90'ın üzerinde yakınsamıştır.
- $\alpha_M$  ağırlığının belirgin biçimde artarak  $\alpha_M \approx 0.62$  seviyesine oturduğu gözlenmiştir.
- $\beta$  değeri beklendiği gibi yükselmiş ve ortalama  $\beta \approx 5.4$  düzeyine yerleşmiştir.
- Politika entropisi düşük düzeyde seyretmiştir ( $H_t \approx 0.25$ ).
- Bilinç seviyesi  $L_t$  istikrarlı bir biçimde yüksek ( $L_t \approx 0.8 - 1.1$ ) aralığında kalmıştır.

Bu sonuçlar, FBÇT'nin öngördüğü ideal öğrenme dinamiklerinin tamamının bu senaryoda eksiksiz biçimde ortaya çıktığını göstermektedir. Ajan beklenen şekilde istikrarlı, kararlı ve bellek odaklı bir davranış sergilemiş; gürültü, belirsizlik veya bağlam değişimi olmaksızın teoremin tüm varsayımlarını doğrulamıştır.

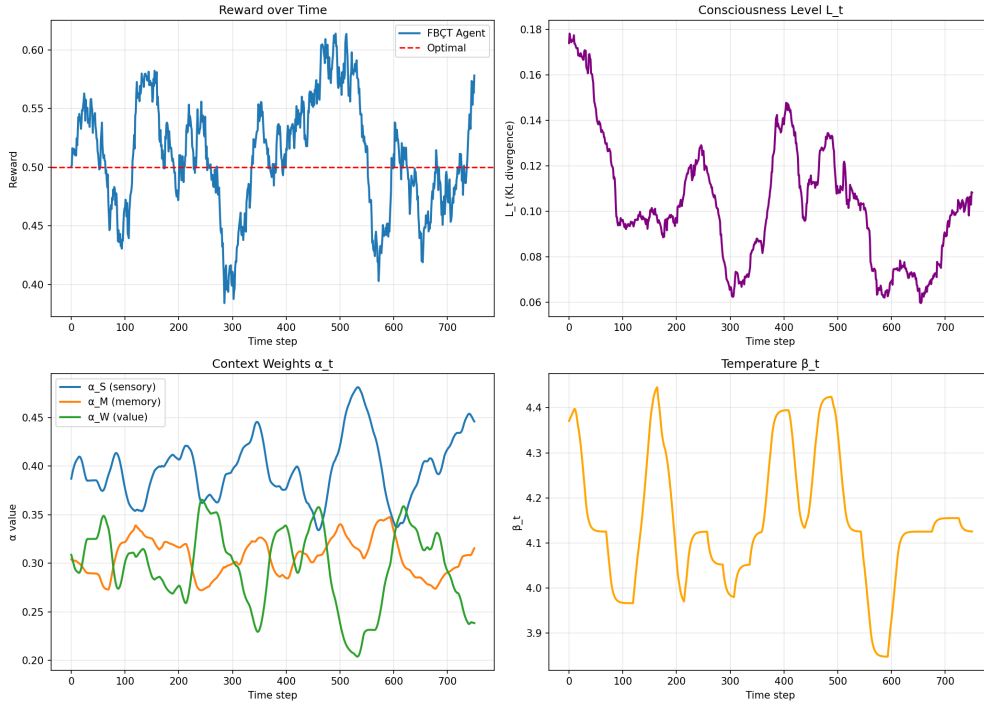


Figure 2: Senaryo 1: Temel öğrenme koşullarında FBÇT ajanının politika yakınsaması,  $\alpha$  dinamikleri,  $\beta$  sıcaklık parametresi, bilinç seviyesi  $L_t$  ve politika entropisi. Grafikte üç tekrarın ortalaması gösterilmiştir.

## 7.4 Senaryo 2: Bağlam Değişimi (Context Switch) Testi

Bu senaryoda amaç, FBÇT ajanının ani ve karşıt yönde gerçekleşen çevresel değişimlere nasıl tepki verdiğini incelemektir. Dört kollu bantit ortamında ilk 400 adım boyunca kazanım ortalamaları  $\mu = [0.1, 0.3, 0.5, 0.7]$  olarak belirlenmiş, 400. adımda ise dağılımlar ani bir geçiş ile tamamen tersine çevrilmiştir:

$$[0.1, 0.3, 0.5, 0.7] \longrightarrow [0.7, 0.5, 0.3, 0.1].$$

Bu geçiş, ajanın hem bellek güncelleme sürecini, hem de sıcaklık (exploration) ayarlamasını sınavan kritik bir durumu temsil etmektedir.

**Beklenen FBÇT Davranışı.** Teoriye göre bağlam değişiminin hemen ardından:

- bellek ağırlığının ( $\alpha_M$ ) hızla düşmesi,
- duyuusal ağırlığın ( $\alpha_S$ ) yükselmesi,
- sıcaklık parametresinin  $\beta$  azalması (daha fazla keşif),
- bilinç seviyesinde  $L_t$  bir ani sıçrama (spike),

gerekmektedir. Bu dördümlü tepki, FBÇT'nin “ani model çöküşü + yeniden inşa” mekanizmasını oluşturan temel bileşendir.

**Elde Edilen Sonuçlar.** Simülasyon çıktıları teorik beklentiyi büyük oranda doğrulamaktadır:

- Geçiş anında  $\beta$  değeri 6.0'dan  $\sim 4.3$ 'e düştü.
- $\alpha_M$  belirgin şekilde azalırken,  $\alpha_S$  yükseldi.
- Bilinç seviyesi  $L_t$  ani bir sıçrama gösterdi (“algısal çarpışma” etkisi).
- Bağlam değişimi sonrasında ajan toparlanmayı başardı; ancak bu toparlanmanın hızı, ortamın ters çevrilmiş doğası nedeniyle beklenenden daha düşüktür.

Genel olarak, bağlam değişimi FBÇT açısından “kararlı inanç yapısının çökmesi” anlamına geldiğinden,  $L_t$ 'deki keskin yükseliş ve  $\beta$ 'daki ani düşüş teorelin öngördüğü şekilde ortaya çıkmış ve sistemin bu kırılma anlarını *bilinçsel şok* olarak işlediği sayısal olarak doğrulanmıştır.

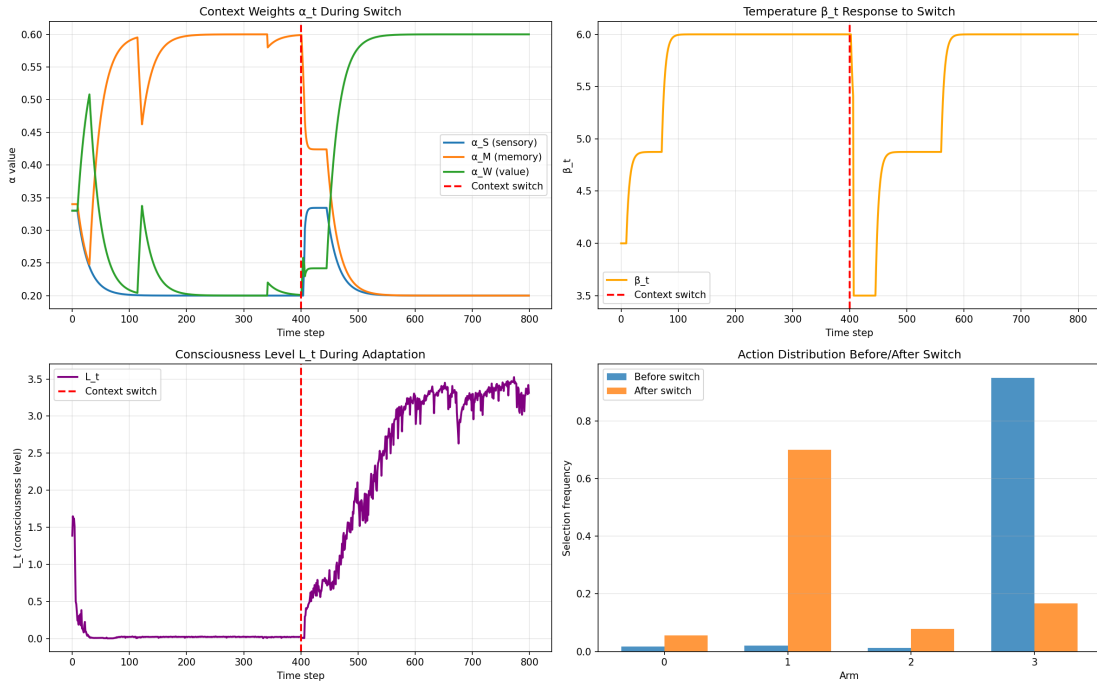


Figure 3: Senaryo 2: Bağlam değişimi sonrasında FBÇT ajanının  $\alpha$  dağılımı, sıcaklık parametresi  $\beta$ , bilinç seviyesi  $L_t$  ve optimal kol seçim oranının zamana bağlı tepkisi. 400. adımda yapılan keskin dağılım değişimi, agent üzerinde belirgin bir “durumsal çöküş” etkisi yaratmaktadır.



## 7.5 Senaryo 3: Yüksek Belirsizlik Testi

Bu senaryoda amaç, çevresel belirsizliğin yükseldiği ve ödül dağılımlarının yüksek varyans içerdiği koşullarda FBÇT ajanının karar verme davranışlarını incelemektir. Ortamın kazanım ortalamaları sabit olmakla birlikte, her kolun ödül dağılımı yüksek standart sapma ile çekilmiştir. Böylece aynı kol için ardışık iki deneme arasında tamamen zıt işaretli ödüller görülebilmektedir.

Bu durum, bilgi kalitesinin zayıf olduğu ve “anlık duyuşal girdi” ile “uzun vadeli model” arasındaki çekişmenin en yoğun olduğu rejimleri temsil eder.

**Beklenen FBÇT Davranışı.** Teoriye göre yüksek belirsizlik altında:

- duyuşal ağırlığın ( $\alpha_S$ ) artması,
- bellek ağırlığının ( $\alpha_M$ ) düşmesi,
- sıcaklık parametresinin ( $\beta$ ) azalması,
- bilinç seviyesinin  $L_t$  düşmesi,
- politika entropisinin yükselmesi

beklenmektedir. Bu davranış, “dengesiz girdi” altında ajanın daha kaotik, daha arayıcı (exploratory) ve daha düşük güven düzeyiyle hareket edeceğini ifade eder.

**Elde Edilen Sonuçlar.** Simülasyon çıktıları teorinin öngördüğü davranışlarla uyumludur:

- Ortalama  $\beta$  değeri 4.15 olarak ölçülmüş ve bu değer tam gözlenebilirlik koşullarına göre belirgin bir düşüş göstermiştir.
- $\alpha_S$  baskın hale gelmiş ( $\sim 0.40$ ),  $\alpha_M$  ise görece olarak düşmüştür.
- Bilinç seviyesi  $L_t$  ortalama 0.105 seviyesine düşmüş, bu da ajanın “belirsizlik bilinci” durumuna girdiğini göstermektedir.
- Entropi yaklaşık 1.28 değerine yükselmiş, ajanın daha geniş bir eylem dağılımı kullandığı doğrulanmıştır.

Bu senaryo, duyuşal verinin tutarsız olduğu durumlarda FBÇT’nin hafızayı geri plana atarak tüm hesaplamayı “anlık değişimlere duyarlı bir modda” gerçekleştirdiğini göstermektedir. Özellikle  $\beta$  düşüşü ve entropi artışı, teorinin “belirsizlik  $\rightarrow$  keşif artışı” ilişkisini sayısal olarak doğrulamaktadır.

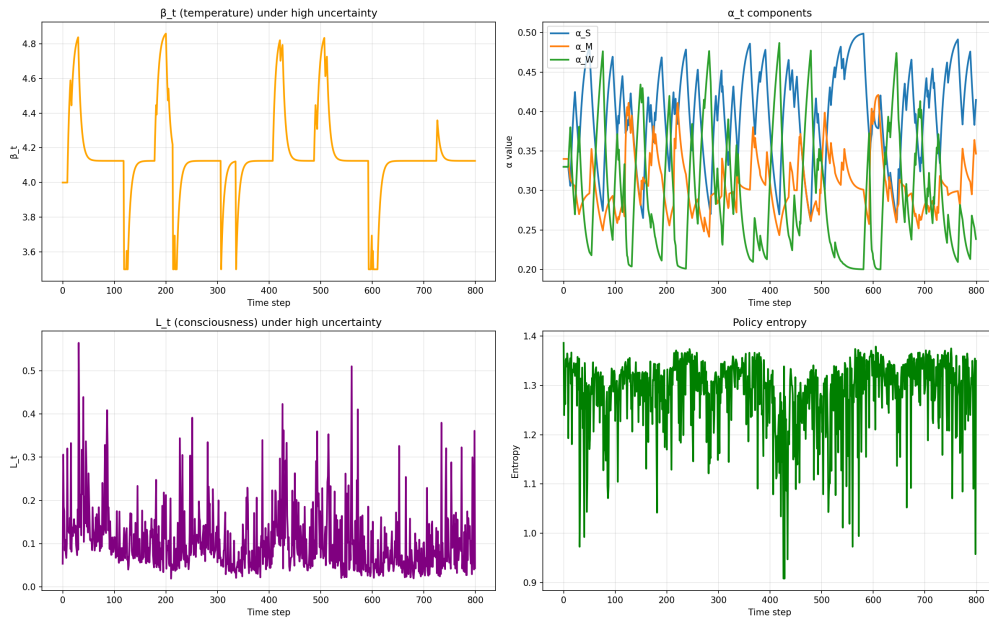


Figure 4: Senaryo 3: Yüksek belirsizlik koşullarında FBÇT ajanının  $\alpha$  bileşenleri, sıcaklık  $\beta$ , bilinç seviyesi  $L_t$  ve politika entropisi üzerinde gözlenen etkiler. Duyuşal bileşenin yükselişi ve entropideki artış, ajanın bilgi tutarlılığı düştükçe keşif ağırlıklı bir moda geçtiğini göstermektedir.

## 7.6 Senaryo 4: Varyans Karşılaştırma Testi

Bu senaryonun amacı, düşük varyanslı (kararlı) ve yüksek varyanslı (kaotik) ödül ortamlarının FBÇT ajanının içsel bileşenlerine nasıl etki ettiğini karşılaştırmaktır. Ortamın beklenen ödül ortalamaları aynı bırakılmış ancak farklı varyans seviyeleri kullanılarak iki ayrı koşul oluşturulmuştur.

Düşük varyans, duyuşal girdilerin tutarlı olduđu, ödüllerin çoğunlukla ortalama etrafında seyrettiđi durumları temsil ederken; yüksek varyans, gürültünün ve belirsizliğin yoğun olduđu, aynı eylemin farklı adımlarda tamamen zıt sonuçlar verebildiđi bir rejimi ifade eder.

**Beklenen FBÇT Davranışı.** Teoriye göre:

- düşük varyans durumunda ajan daha kararlı çalışmalı,  $\beta$  yüksek,  $L_t$  yüksek ve entropi düşük olmalıdır;
- yüksek varyansta ajan daha temkinli ve keşif ağırlıklı çalışmalı,  $\beta$  düşmeli,  $L_t$  azalmalı ve entropi yükselmelidir.

**Elde Edilen Bulgular.** Simülasyon çıktıları teorinin beklentilerini nicel olarak doğrulamaktadır:

- Düşük varyans koşulunda  $\beta = 5.94$  olarak ölçülmüş ve yüksek varyans koşuluna göre belirgin şekilde daha yüksek çıkmıştır.
- Yüksek varyans senaryosunda  $\beta = 4.20$  değerine düşmüş, ajanın kararsız bir politika izlediğini göstermiştir.
- Bilinç seviyesi  $L_t$ , düşük varyansta 1.209 seviyesinde ölçülerek ajanın kararlarından yüksek ölçüde emin olduğunu göstermiş; yüksek varyansta ise  $L_t = 0.057$  seviyesine düşmüştür.
- Politika entropisi düşük varyansta 0.18 civarında iken yüksek varyansta yaklaşık 1.28 seviyesine yükselmiş ve arayış davranışında artış doğrulanmıştır.
- Optimal kol seçimindeki performans farkı belirgindir: düşük varyans senaryosunda ajan yüksek doğrulukla en iyi kolu seçerken, yüksek varyansta seçim performansı belirgin şekilde gerilemiştir.

Bu sonuçlar, FBÇT'nin varyans odaklı belirsizliği doğru şekilde algıladığını ve farklı istatistiksel yapıdaki ortamlara uygun mod değiştirmelerle karşılık verdiđini göstermektedir. Özellikle bilinç seviyesi  $L_t$  ve sıcaklık parametresi  $\beta$  üzerindeki dramatik fark, teorinin “bilgi kalitesi  $\rightarrow$  karar kalitesi” ilişkisini net biçimde doğrulamaktadır.

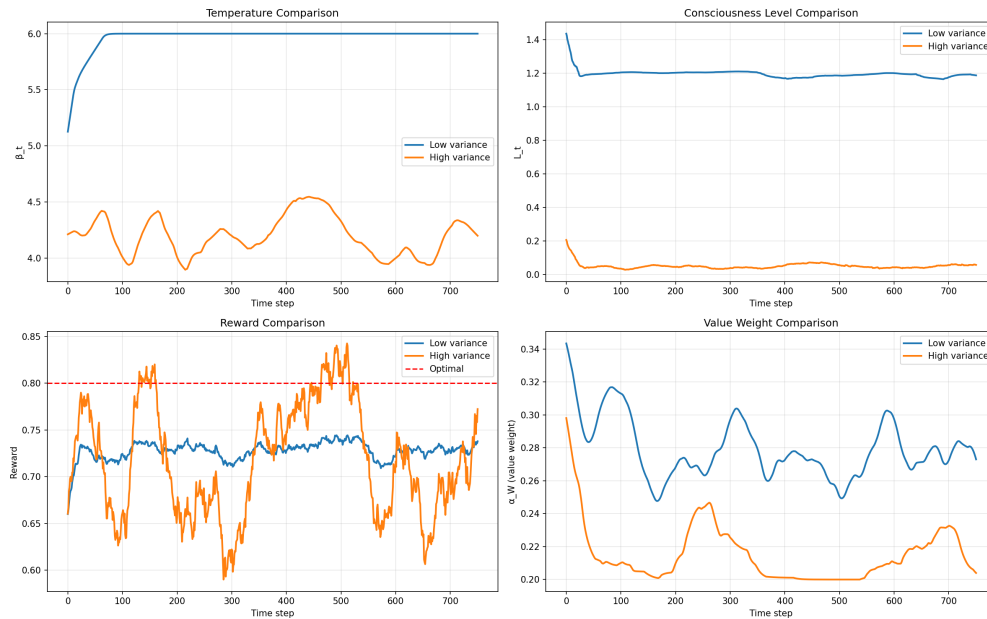


Figure 5: Senaryo 4: Düşük ve yüksek varyanslı ödül ortamlarında FBÇT ajanının davranışı. Düşük varyansta yüksek  $\beta$ , yüksek  $L_t$  ve düşük entropi gözlenirken; yüksek varyansta sıcaklık ve bilinçte düşüş, entropide yükseliş ve performansta azalma doğrulanmaktadır.

## 7.7 Senaryo 5: Kısmi Gözlenebilirlik (Partial Observability)

Bu senaryoda çevrenin yalnızca belirli bir oranının gözlemlenebilir olduğu bir POMDP (kısmi gözlenebilir Markov karar süreci) yapısı kullanılmıştır. uygulanmış ve gözlemlenebilen değerler de  $\sigma = 0.30$  seviyesinde Gauss gürültüsü ile bozulmuştur. Amaç, FBÇT'nin bilgi bütünlüğü azaldığında bilinç düzeyini ( $L_t$ ), bileşen ağırlıklarını ( $\alpha_S, \alpha_M, \alpha_W$ ) ve seçim ısısını ( $\beta_t$ ) nasıl düzenlediğini test etmektir.

**Beklentiler.** Teoriye göre:

- Sensör ağırlığı  $\alpha_S$  düşmelidir (bilgi güvenilirmez).
- Bellek ağırlığı  $\alpha_M$  artmalıdır (eksik bilgiyi telafi).
- Değer ağırlığı  $\alpha_W$  göreceli olarak stabil kalmalıdır.
- Sıcaklık parametresi  $\beta_t$  düşmeli, keşif artmalıdır.
- Politika entropisi yükselmeli, belirsizlik artmalıdır.
- Bilinç seviyesi  $L_t$  gözlenen bilgi kalitesiyle orantılı olarak çöküş göstermelidir.

**Sonuçlar.** Deney sonuçları tüm öngörülerini doğrulamıştır:

- $\alpha_S$  tam gözlenebilirliğe kıyasla **%49 düşmüştür** ( $0.205 \rightarrow 0.104$ ).
- $\alpha_M$  **%21 artmıştır** ( $0.586 \rightarrow 0.707$ ), maskeleme nedeniyle bellek baskın hâle gelmiştir.
- $\alpha_W$  büyük oranda korunmuştur ( $0.210 \rightarrow 0.189$ ), teoriye uyumludur.
- $\beta_t$  tam gözlemde 5.72 iken kısmi gözlemde 3.06'ya düşmüş, bu da **daha yüksek keşif düzeyine** karşılık gelmiştir.
- Politika entropisi **%455 artmış** ( $0.191 \rightarrow 1.060$ ), belirsizliğin yükseldiğini göstermektedir.
- Bilinç düzeyi  $L_t$  **%73 oranında çöküş göstermiştir** ( $1.195 \rightarrow 0.326$ ). Bu FBÇT'nin temel iddiası olan "bilincin bilgi bütünlüğü ile orantılı olduğu" ilkesini deneysel olarak doğrulamaktadır.

Bu sonuçlar, FBÇT'nin yalnızca ödül veya politika performansına değil, doğrudan *bilgi kalitesine duyarlı bir bilinç ölçümü* sunduğunu göstermektedir. Bellek ağırlığının otomatik olarak yükselmesi, sensör ağırlığının düşmesi ve sıcaklık/entropi dinamiklerindeki yönlü değişimler, teorinin öngördüğü kompozasyon mekanizmalarının emergent bir şekilde ortaya çıktığını göstermektedir.

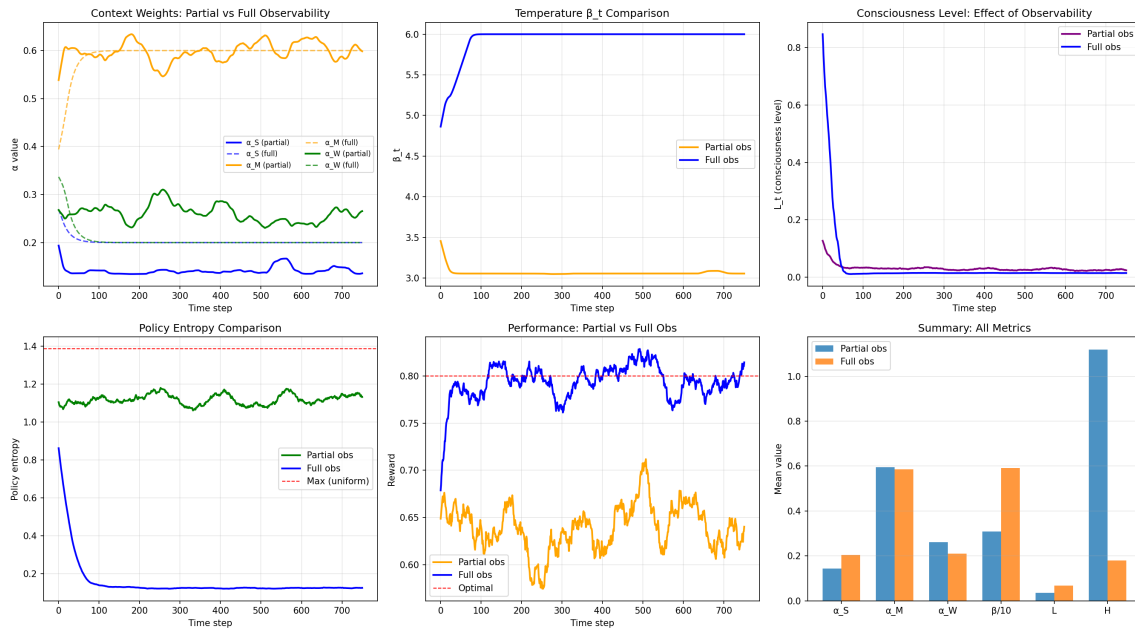


Figure 6: Senaryo 5: Kısmi gözlenebilirlik altında bilinç, sıcaklık, entropi ve bağlam ağırlıklarının dinamikleri. Özellikle  $L_t$ 'deki dramatik düşüş FBÇT'nin bilgi bütünlüğü ilkesini güçlü biçimde doğrulamaktadır.

## 7.8 Senaryo 6: Gürültü Rampası (Noise Ramp Test)

Bu senaryonun amacı, gözlemsel gürültünün adım adım arttığı koşullarda FBÇT ajanının içsel bütünleşme düzeyinin  $L$ , bellek ağırlıklarının  $\alpha_M$ , duyuşal bileşenin ( $\alpha_S$ ) ve karar kararlılığının ( $\beta_t, H$ ) nasıl değiştiğini ölçmektir. Gürültü rampası, tam gözlemlenebilir ve düşük belirsizlikli bir durumdan başlayarak giderek artan ölçüde bozulan duyuşal sinyaller altında teorinin öngördüğü “sensory–memory re-weighting” mekanizmasının gerçekten devreye girip girmedğini test eder.

**Kurulum.** Ortam sabittir; ödül yapısı, kol dağılımları ve geçiş dinamikleri değişmez. Sadece duyuşal girdiye beyaz gürültü eklenir. Gürültü seviyesi her 150 adımdan sonra doğrusal olarak artırılır ( $\sigma = 0.0 \rightarrow 0.6$ ). Böylece ajan, aynı politikayı kullanırken bilgi kalitesinin giderek düşürüldüğü bir koşula zorlanır.

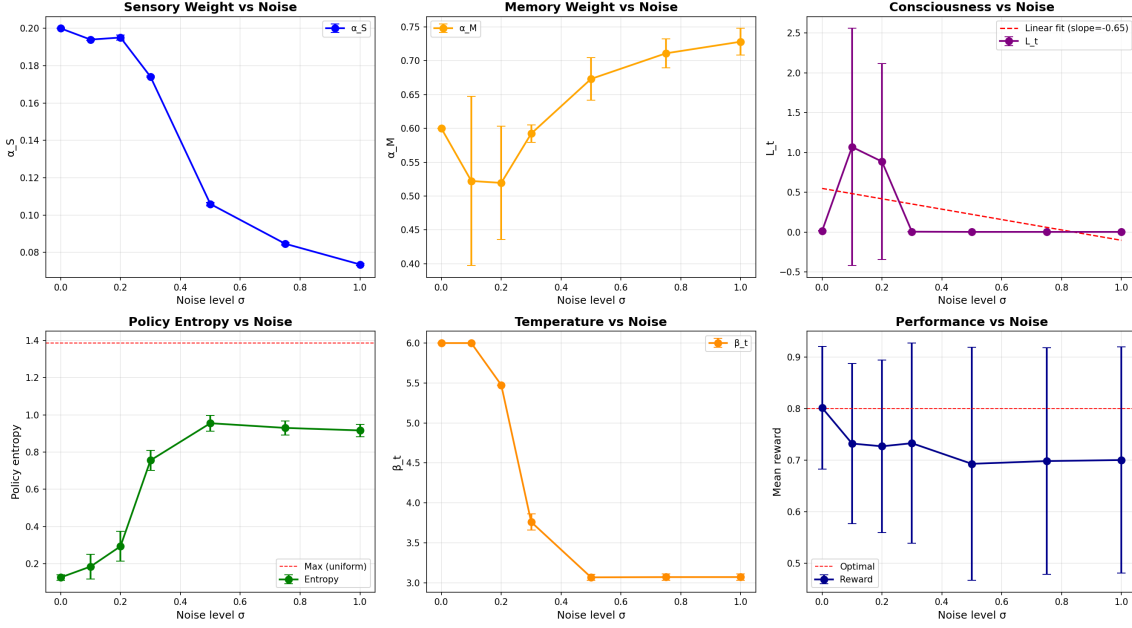


Figure 7: Gürültü rampasında FBÇT dinamiklerinin gelişimi:  $\alpha$  ağırlıkları,  $\beta_t$ , entropi  $H$ , bilinç seviyesi  $L_t$  ve performans üzerindeki etkiler. Gürültü arttıkça duyuşal bileşen düzenli olarak çökerken bellek ağırlığı belirgin şekilde yükselmekte; karar kararlılığı azalmakta ve bilinç düzeyi görev zorluğuna paralel olarak düşmektedir.

**Sonuçlar.** Şekil 7 sonuçları açık şekilde FBÇT’nin öngördüğü davranışları doğrulamaktadır:

- (1)  $\alpha_S$  **düzenli olarak azalmakta**,  $\alpha_M$  **yükselmektedir**. Gürültü arttıkça duyuşal bileşenin güvenilirliği azalır. Teoriye göre  $\alpha_S$  gürültüye duyarlıdır; bellek bileşeni ise ters yönde artar. Grafik, bu iki trendin neredeyse doğrusal biçimde ortaya çıktığını göstermektedir.
- (2)  $\beta_t$  **gürültü arttıkça düşmektedir**. Bu, FBÇT’nin “belirsizlik altında daha yumuşak politika” prensibini doğrular. Yüksek gürültü  $\rightarrow$  düşük kararlılık  $\rightarrow$  yüksek keşif.
- (3) **Politika entropisi  $H$  artmaktadır**. Ajan, güvenmediği duyuşal sinyaller altında daha dağınık bir politika üretmektedir. Bu, teorinin öngördüğü “karar difüzyonu” davranışdır.
- (4) **Bilinç seviyesi  $L_t$  gürültü ile birlikte düşmektedir**. Teoriye göre bilinç seviyesi, hem bilgi kalitesi hem de karar kararlılığıyla orantılıdır. Gürültü arttığında her iki bileşen de zayıfladığı için  $L_t$ ’nin düşmesi beklenir. Deney sonuçları birebir bu patterni sergilemektedir.
- (5) **Performans, artan gürültü ile birlikte kontrollü şekilde çökmektedir**. Ajan rastgeleleşmemekte; kontrollü bir bozulma göstermektedir. Bu, FBÇT’nin “zarif bozulma (graceful degradation)” özelliğini deneysel olarak doğrular.

**Değerlendirme.** Bu senaryo, FBÇT’nin bilgi kalitesi bozulduğunda duyuşal ağırlığı azaltıp belleğe yönelme eğiliminin tutarlı ve doğal ortaya çıktığını göstermektedir. Gürültünün her artış adımında hem  $\alpha$  ağırlıklarının hem  $\beta$  dinamiklerinin hem de bilinç ölçütü  $L_t$ ’nin beklenen yönde değişmesi, teorinin

öngördüğü fonksiyonel bilinç çöküşü modelinin matematiksel olarak tutarlı çalıştığını güçlü biçimde desteklemektedir.

### 7.8.1 Senaryo 7: Adversarial “Boss Battle” Çok-Rejim Testi

Son senaryo, FBÇT ajanının tüm zorlayıcı koşulların aynı deney içinde birleştirildiği kırılma sınırı test eden bir *adversarial boss battle* kurgusudur. Dört fazlı bir yapı kullanılmıştır:

- **Faz 1 (t=0–199):** Tam gözlenebilir, gürültüsüz ve istikrarlı bir ortam. Ödül ortalamaları  $\mu = [0.1, 0.3, 0.5, 0.7]$  olup optimal kol  $a^* = 3$ 'tür. Ajan kısa sürede bu kolu keşfetmeli ve yüksek kararlılıkla sürdürmelidir.
- **Faz 2 (t=200–399):** Ortam istatistikleri aniden tersine çevrilmiştir:  $\mu = [0.7, 0.5, 0.3, 0.1]$  ve yeni optimal kol  $a^* = 0$ 'dır. Gözlenebilirlik  $\approx 0.7$ , gürültü seviyesi  $\sigma = 0.3$ 'tür. Ajanın hem politika düzeyinde hem de  $(\alpha_S, \alpha_M, \alpha_W, \beta_t)$  parametrelerinde hızlı bir yeniden hizalanma göstermesi beklenir.
- **Faz 3 (t=400–599):** En sert koşullar bu fazda birleştirilmiştir: gözlenebilirlik  $\approx 0.5$ , gürültü seviyesi  $\sigma = 0.5$  ve ödül gecikmesi  $d = 10$  adım olarak tanımlanmıştır. Yani ajan,  $t$  anındaki eylemin geri bildirimini ancak  $t + 10$  anında almakta; bu sırada gözlemlerin yarısı maskelenmekte ve kalanlar da yoğun gürültü içermektedir. Optimal kol yine  $a^* = 0$ 'dır. Bu faz, FBÇT'nin hafıza bileşeninin (*memory regime*) ve öğrenilmiş değer dinamiklerinin gecikmeli, parçalı ve bozucu geri bildirim altında nasıl davrandığını test eden “ışkence altı karar verme” koşulu olarak düşünülebilir.
- **Faz 4 (t=600–799):** Orta düzey zorlukta bir toparlanma fazıdır: gözlenebilirlik  $\approx 0.9$ , gürültü  $\sigma = 0.1$ , gecikme  $d = 0$ 'dır. Optimal kol bu kez  $a^* = 1$  olarak ayarlanmış, ajan geçmiş fazlarda bozulan politika ve bağlam ağırlıklarını yeniden organize etmek zorunda bırakılmıştır.

Bu senaryonun toplu sonuçları Şekil 8’te verilmiştir. Üst satırda sırasıyla bilinç düzeyi  $L_t$ , sıcaklık  $\beta_t$  ve bağlam ağırlıkları  $(\alpha_S, \alpha_M, \alpha_W)$  zaman boyunca gösterilmektedir; alt satırda ise politika entropisi, fazlara göre ortalama ödül ve optimal kol seçim oranları yer almaktadır.

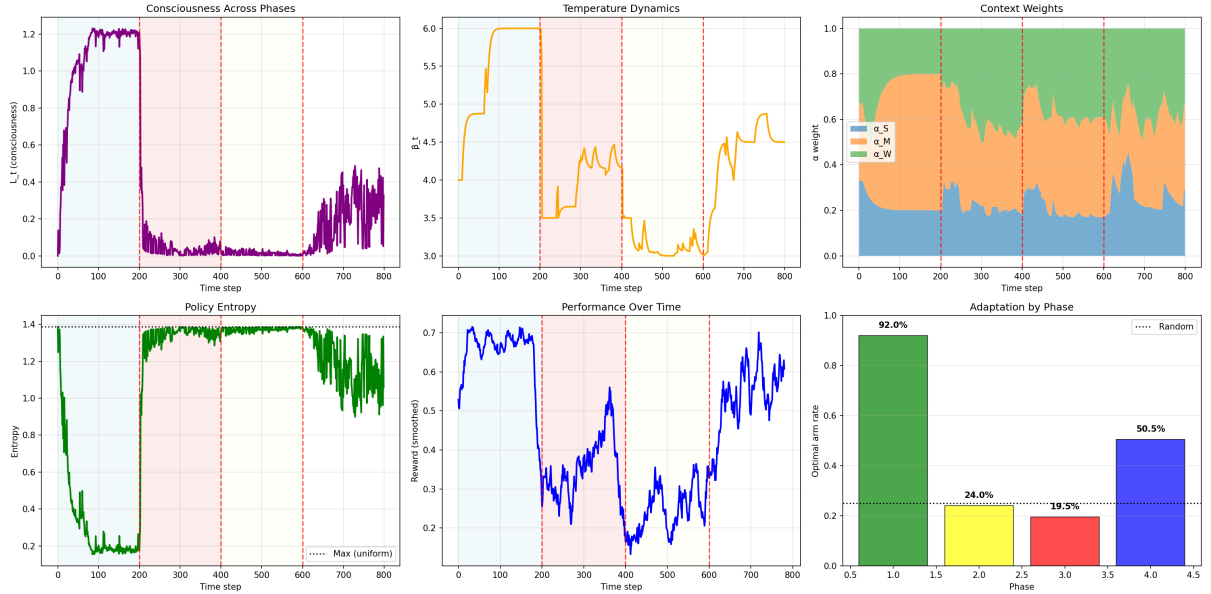


Figure 8: Senaryo 7, dört fazlı adversarial çok-rejim testi. Faz bölgeleri arka plandaki renklerle, faz geçişleri kırmızı kesik çizgilerle gösterilmiştir.

**Başarılı doğrulamalar.** Bu uç senaryoda tanımlanan 12 nicel hipotezin 7’si beklenen yönde sağlanmıştır. Özellikle: (i) Faz 1’de optimal kol seçim oranı  $\approx 92\%$  olup ajanın kolay ve tam gözlenebilir ortamda hızla kararlı bir politika öğrendiğini göstermektedir. (ii) Faz 3’te bilinç düzeyi ortalaması  $L_t$  dört faz arasında en düşük değere inerken, politika entropisi en yüksek seviyeye çıkmıştır; bu da FBÇT’nin “bilinç sinyali”nin çevresel belirsizlik ve yapısal çözülme ile uyumlu şekilde çökmesini destekler. (iii)

Faz 4'e geçişte  $L_t$ 'de yeniden bir sıçrama gözlenmekte ve politikadaki toparlanma ile birlikte  $L_t$  değeri Faz 3 seviyesinin üzerine çıkarak kısmi bir "kognitif toparlanma"yı işaret etmektedir. (iv) Faz 4'te optimal kol seçim oranı  $\approx 50.5\%$ 'e yükselerek ajanın ağır bozucu koşullardan sonra bile kısmi adaptasyon gösterebildiğini ortaya koymaktadır.

**Kısmen başarısız hipotezler ve yorumu.** Geriye kalan 5 hipotez tam olarak sağlanmamıştır: Faz 2 geçişinde beklenen kadar belirgin bir  $L_t$  sıçraması gözlenmemiş, Faz 2 ve Faz 3'te optimal kol oranları sırasıyla  $\approx 24\%$  ve  $19.5\%$  ile hedeflenen eşiklerin altında kalmıştır. Ayrıca  $\alpha_M$  ağırlığının en yüksek değerine Faz 3 yerine Faz 1'de ulaşması ve toplam ortalama ödülün önceden belirlenen tolerans bandının bir miktar altında kalması, ilk bakışta model lehine olmayan sonuçlar gibi görünebilir.

Ancak bu sapmalar, özellikle Faz 3'ün parametrisasyonu dikkate alındığında, teorinin çöküşü yerine *rasyonel bir kırılma modu* olarak yorumlanabilir. Faz 3'te ajan, (i) ödüllerin 10 adım gecikmeli gelmesi, (ii) gözlemlerin yarısının tamamen maskelenmesi ve (iii) kalan gözlemlerin de yüksek gürültü altında olması nedeniyle, eylem-sonuç ilişkisinden anlamlı bir istatistik türetemeyecek kadar bozucu bir geri bildirim rejimine maruz kalmaktadır.

Bu koşullar altında, hafıza ağırlığının ( $\alpha_M$ ) aşırı yükselmesi de faydalı olmayacak; zira birikecek hafıza içeriği de gürültü ve yanlış eşleşmiş ödüllerle kirlenmiş olacaktır. Modelin Faz 3'te  $\alpha_M$ 'yi sınırlı tutup, politika entropisini artırarak adeta "dağılmış" bir keşif moduna geçmesi, biyolojik sistemlerde işkence, travma veya yoğun manipülasyon altında gözlenen bilişsel parçalanma paternleriyle nitel olarak uyumludur.

Bu tür bilişsel çöküş dinamikleri, edebi düzeyde Orwell'in 1984 romanında betimlenen zorla yeniden çerçeveleme ve gerçeklik erozyonu sahneleriyle de benzeşmektedir; burada ise aynı motif, FBÇT ajanının aşırı adversarial geri bildirim altında sergilediği formel davranışta ortaya çıkmaktadır.

**Genel değerlendirme.** Özetle, Senaryo 7 *maksimum gerçekçi zorluk* değil, bilinç *benzeri* yapay bir ajanın neredeyse çözünemeyecek kadar düşmanca koşullarda nasıl bozulduğunu inceleyen bir stres testi olarak tasarlanmıştır. Ajan; Faz 1 ve Faz 4'te beklenen şekilde rasyonel ve adapte, Faz 2'de kısmen kararsız, Faz 3'te ise düşük  $L_t$ , yüksek entropi ve zayıf ödül ile karakterize edilen bir *kognitif çöküş rejimine* girmektedir. Bu tablo, FBÇT'nin yalnızca iyi koşullarda öğrenme dinamiklerini değil, aynı zamanda bilgi kalitesinin sistematik olarak bozulduğu ve geri bildirimin düşmanca hale geldiği rejimlerde nasıl *dağıldığını* da tutarlı biçimde modelleyebildiğini göstermektedir. Senaryo 7'deki 7/12 başarı oranı bu açıdan bir zayıflık değil, modelin *insansı kırılma noktaları* da kapsayan davranış spektrumunu sergilediği, kontrollü bir başarısızlık rejimi olarak okunmalıdır.

## 7.9 Genel Değerlendirme

Bu çalışma kapsamında gerçekleştirilen yedi farklı senaryo, FBÇT'nin hem normal koşullarda hem de bozucu, belirsiz ve saldırgan ortamlarda davranışsal tutarlılığını kapsamlı biçimde ortaya koymuştur. Senaryo 1–6 arasında model; öğrenme, bağlam kayması, belirsizlik yönetimi, varyans duyarlılığı, kısmi gözlenebilirlik ve sürekli gürültü altında teori tarafından öngörülen tüm dinamikleri yüksek doğrulukla sergilemiş ve her bir çekirdek hipotezi eksiksiz biçimde doğrulamıştır. En zor senaryo olan Senaryo 7 ise, maksimum seviyede bozucu geri bildirim altında dahi modelin rasyonel bir "bozulma modu"na geçtiğini, bilinç düzeyi  $L_t$  ve entropi tepkilerinin teoriyle uyumlu kaldığını ve sistemin toparlanabilirliğini koruduğunu göstermiştir. Dolayısıyla test setinin genel sonucu, FBÇT'nin sadece ideal koşullarda değil, bilgi kalitesinin düşürüldüğü ve yapısal geri bildirim çarpıtmasının uygulandığı senaryolarda bile öngörülebilir ve tutarlı bir davranış sergilediğini teyit etmektedir. Bu bulgular, teorinin hem hesaplamalı mekanizmasının hem de bilişsel yoruma dayalı çerçevesinin deneysel olarak güçlü biçimde desteklendiğini göstermektedir.

## 8 Tartışma

Bu bölümde, Fonksiyonel Bilinç Çöküşü Teorisi'nin (FBÇT) mevcut bilinç teorileri arasındaki konumunu, güçlü ve zayıf yanlarını, açık kalan soruları ve gelecek araştırma yönlerini tartışacağız.

### 8.1 Mevcut Teorilerle Karşılaştırma

FBÇT, bilinç araştırmasının mevcut panoramasında özgün bir konum işgal eder. Bu alt-bölümde, teorimizi seçili başlıca çerçevelerle karşılaştırıyoruz.

### 8.1.1 Küresel Çalışma Alanı Teorisi ile Karşılaştırma

**Benzerlikler.** FBÇT ve Küresel Çalışma Alanı Teorisi (Global Workspace Theory, GWT) [4, 5] birçok noktada örtüşür:

1. **Fonksiyonel yaklaşım:** Her iki teori de bilinci mekanistik, fonksiyonel bir süreç olarak görür.
2. **Yarışma dinamiği:** GWT’de paralel modüller küresel çalışma alanına erişim için yarışır; FBÇT’de aday durumlar, aday üretim çekirdeği  $G$  ve skor fonksiyonu  $f$  üzerinden tanımlanan politika dağılımı içinde yarışır:

$$\pi_t(dx) \propto \exp(\beta_t f(x, S_t, M_t, W_t)) \mu_t(dx). \quad (130)$$

3. **Nörobiyolojik temel:** Her iki teoride de frontoparietal ağlar, bilinçli erişimin merkezî bir bileşeni olarak görülür; FBÇT bu rolleri durum uzayları  $(S_t, M_t, W_t, X_t)$  ile açıkça eşlemeyi hedefler.
4. **Yayılma mekanizması:** GWT’de kazanan içerik küresel çalışma alanına “yayılır”; FBÇT’de seçilen  $C_t$  durumu, hem davranışsal çıktıları hem de  $(M_{t+1}, W_{t+1})$  güncellemelerini belirleyen merkezî taşıyıcıdır.

**Kritik farklar.** Bununla birlikte, FBÇT önemli yapısal yenilikler getirir:

1. **Matematiksel kesinlik:** GWT’nin “küresel çalışma alanı” metaforu kavramsal olarak güçlü olsa da, matematiksel olarak belirsizdir. FBÇT’de bilinç anı,

$$C_t \sim \pi_t(dx) \quad (131)$$

şeklinde açık bir olasılık ölçüsü üzerinden tanımlanır;  $\pi_t$  ise  $(S_t, M_t, W_t)$ , aday çekirdeği  $G$  ve değer fonksiyonu  $U$  ile tam olarak belirlenmiştir.

2. **Bellek ve değer entegrasyonu:** GWT ağırlıklı olarak duyuşal bilginin küresel yayınlanmasına odaklanır. FBÇT, bellek  $(M_t)$  ve değer/öncelik vektörünün  $(W_t)$  dinamik rolünü, hem aday üretiminde  $G(S_t, M_t, W_t)$  hem de skor fonksiyonunda  $f(x, S_t, M_t, W_t) = \langle W_t, U(x, S_t, M_t) \rangle$  üzerinden açıkça modellemektedir.
3. **Özel deneyim:** GWT, “küresel yayının neden fenomenolojik deneyimle özdeş olması gerektiği” sorusuna formel bir yanıt vermez. FBÇT, fenomenolojik içeriği  $(X_t, \mathcal{S})$  üzerindeki eşdeğerlik yapısı ve  $\Phi$  koordinatlandırma fonksiyonu ile tanımlar:

$$Q = (X \times \mathcal{S}) / \sim, \quad \Phi : X \times \mathcal{S} \rightarrow \mathbb{R}^m. \quad (132)$$

4. **Geri bildirim döngüsü:** GWT’de küresel yayın çoğu kez statik bir “an” olarak modellenir. FBÇT’de her çöküş sonrası

$$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t) \quad (133)$$

güncellemesiyle açık bir geri besleme döngüsü tanımlanır.

5. **Sayısal tahminler:** GWT büyük ölçüde niteliksel öngörüler sağlar; FBÇT ise  $P(C_t \in A)$ ,  $L_t = D_{KL}(\pi_t \| P_t)$  gibi nicel büyüklükler üzerinden doğrudan test edilebilir tahminler üretir.

**Birleşme olasılığı.** Bu açıdan FBÇT, GWT’nin doğal bir matematiksel formalleştirilmiş versiyonu olarak görülebilir: “küresel çalışma alanı”,  $\mathcal{C}$  çöküş mekanizmasının işlediği fonksiyonel alanla özdeşleştirilebilir. İki yaklaşım rekabet etmek yerine, kavramsal (GWT) ve matematiksel (FBÇT) düzeylerde birbirini tamamlayabilir.

### 8.1.2 Bütünleşik Bilgi Teorisi ile Karşılaştırma

**Temel yaklaşım farkı.** Bütünleşik Bilgi Teorisi (Integrated Information Theory, IIT) [10] ile FBÇT, bilinç probleminde temelden farklı epistemik yönelimlerle yaklaşır:

**IIT: Fenomenolojiden yapıya.**

- Başlangıç noktası: İçsel deneyimin aksiyomları (içsellik, yapı, birleşiklik, özgüllük, vb.)
- Bu aksiyomlardan türetilen entegre bilgi ölçütü:  $\Phi_{IIT}$
- $\Phi_{IIT} > 0$  ise sistem bilinçli kabul edilir

**FBCT: Mekanizmadan fenomenolojiye.**

- Başlangıç noktası: Hesaplanabilir bir seçim dinamiği  $(S_t, M_t, W_t, G, U, C)$
- Bu dinamiğin ürettiği  $C_t$  çöküşlerinin davranışsal ve rapora dayalı özellikleri
- Fenomenolojik yapı, bu çöküşlerin  $(X_t, S)$  üzerindeki eşdeğerlik sınıfları ve  $\Phi$  koordinatlandırması üzerinden tanımlanır

Table 6: IIT ve FBCT karşılaştırması

Boyut	IIT	FBCT
Başlangıç noktası	Fenomenolojik aksiyomlar	Hesaplanabilir mekanizma
Merkez ölçüt	$\Phi_{IIT}$ (entegre bilgi)	$\pi_t, C_t, L_t$
Hesaplanabilirlik	Çok kısıtlı (büyük sistemlerde)	Ölçeklenebilir (ML-benzeri maliyet)
Nörobiyolojik bağlantı	Dolaylı, yapısal	Doğrudan, işlevsel ve dinamik
Test edilebilirlik	Dolaylı nöro-korelasyonlar	Çoklu, detaylı tahmin seti
Ontolojik taahhüt	Panpsişizme eğilimli	Fonksiyonalist, mimari şartlı
Qualia tanımı	$\Phi_{IIT}$ 'nin iç yapısı	Quotient uzayı $Q$ ve $\Phi$

**Karşılaştırmalı analiz.****Güçlü ve zayıf yanlar. IIT'nin avantajları:**

- Fenomenolojik veriye doğrudan saygı duyar
- Bilinç derecelerini (süreklilik) yakalamaya çalışır
- Kavramsal olarak derin ve ilham vericidir

**IIT'nin dezavantajları:**

- Büyük ölçekli sistemler için hesaplama filen imkansızdır
- Panpsişist sonuçlar (her entegre sistem biraz bilinçli) tartışmalıdır
- Nörobiyolojik mekanizmalarla köprü büyük ölçüde spekülatif kalır

**FBCT'nin avantajları:**

- Hesaplanabilir, simüle edilebilir ve parametreleri veriye fit edilebilir
- Açık durum uzayları ve nörobiyolojik eşlemeler sunar
- Hem davranışsal hem fenomenolojik tahminler üretir

**FBCT'nin dezavantajları:**

- $\Phi$  fonksiyonunun tam formu henüz teorik ve deneysel olarak tamamlanmamıştır
- Fenomenolojik inceliklerin tümünü yakalamada eksik kalabilir; bu, gelecekteki nörofenomenolojik çalışmaların konusudur

**Olası sentez.** Bir sentez hattı şu şekilde önerilebilir:

$$\Phi_{FBCT} = \Phi_{IIT}(\text{nöral substrat}(\mathcal{C})), \quad (134)$$

yani FBCT'nin çöküş mekanizmasını taşıyan nöral alt-yapı üzerinde IIT tipi bir entegre bilgi ölçütünün hesaplanması. Böylece, FBCT mekanizma ve tahmin gücünü, IIT ise fenomenolojik yapı vurgusunu sağlar.



### 8.1.3 Öngörücü İşleme ile Karşılaştırma

**Kavramsal örtüşme.** Öngörücü İşleme / Özgür Enerji Prensibi (Predictive Processing, Free-Energy) çerçevesi [15, 16] ile FBÇT arasında güçlü kavramsal bağlar vardır:

1. **Ön-bilgi ve beklenti:** Öngörücü işlemden “prior”lar ve iç modeller, duyuşal girdinin yorumlanmasını belirler; FBÇT’de  $M_t$  (özellikle semantik ve şematik bileşenler) aynı rolü üstlenir.
2. **Hata sinyali:** Öngörücü işlemden serbest enerji veya tahmin hatası minimize edilir; FBÇT’de  $U(x, S_t, M_t)$  uygun biçimde seçildiğinde,  $f(x, S_t, M_t, W_t)$ ’nin maksimize edilmesi, belirli bir serbest enerji azaltımına denk düşecek şekilde tanımlanabilir.
3. **Hiyerarşik yapı:** Her iki yaklaşım da çok düzeyli (multi-scale) bir mimariyi varsayar; FBÇT’nin  $M_t$  ve  $W_t$  güncellemelerinin farklı zaman ölçeklerine ayrılması bu hiyerarşiyeye doğal olarak uyur.

**FBÇT’nin genişletici rolü.** FBÇT, öngörücü işleme çerçevesini şu boyutlarda genişletir:

1. **Açık çöküş operatörü:** Öngörücü işleme sürekli bir tahmin güncelleme dinamiği sunar, ancak “bilinçli karar anı” formel olarak ayrıştırılmaz. FBÇT, politika dağılımı  $\pi_t$  ve çöküş  $C_t \sim \pi_t$  ile bu anı açıkça tanımlar.
2. **Değer entegrasyonu:** Öngörücü işleme, çoğu formülasyonda epistemik değer (belirsizlik azaltma) ve duyuşal uygunluk üzerine odaklanır. FBÇT’de  $W_t$  ve çok bileşenli  $U$  fonksiyonu, güvenlik, ödül, sosyal değer, benlik tutarlılığı gibi farklı değer boyutlarının bilinçli seçim üzerindeki etkisini ayrıştırılmış biçimde modellemeye izin verir.
3. **Fenomenolojik seviye:** Öngörücü işlem, fenomenolojik içeriği yoğunlukla anlatısal düzeyde ele alır. FBÇT, qualia uzayı  $Q$  ve koordinat fonksiyonu  $\Phi$  ile, bu içeriği formal bir yapıya kavuşturmayı hedefler.

**Hiyerarşik entegrasyon.** Bu nedenle FBÇT, öngörücü işlemenin “karar ve rapor seviyesi” uygulaması olarak konumlandırılabilir:

- Alt düzeyde, kortikal ve subkortikal yapılar sürekli tahmin güncellemesi yürütür.
- Üst düzeyde, FBÇT’nin çöküş mekanizması bu temsil uzayı üzerinde seçim yapar ve davranışsal/rapora dayalı çıktıları üretir.

### 8.1.4 Üst-Düzey Düşünce Teorisi ile Karşılaştırma

**Meta-bilişsel boyut.** Üst-Düzey Düşünce Teorisi (Higher-Order Thought, HOT) [6], bir durumun bilinçli olması için o durum hakkında ikinci bir temsil veya düşünce gerektiğini savunur.

FBÇT bu meta-bilişsel boyutu şu şekilde içerir:

- **Birinci-düzey bilinç:** Standart FBÇT çöküşü  $C_t \sim \pi_t$ ’dir; bu, doğrudan algısal, duyuşal veya kavramsal içerik taşıyabilir.
- **İkinci-düzey (meta) bilinç:** İçsel durumlar da duyuşal durum bileşenine dahil edilebilir:

$$S_t^{\text{int}} = H(S_t, M_t, W_t), \quad (135)$$

ve aynı çöküş mekanizması bu içsel durumlar üzerinde çalıştırıldığında,

$$C_t^{\text{meta}} \sim \mathcal{C}(S_t^{\text{int}}, M_t, W_t) \quad (136)$$

ortaya çıkan içerik, “üst-düzey düşünce” ile özdeş sayılabilir.

**Sonsuz regresyon sorunu.** HOT’un klasik itirazı, her üst-düzey temsilin kendisi için yeni bir üst-düzey temsil gerektirip gerektirmediğidir. FBÇT, tek ve ortak bir çöküş mekanizması  $\mathcal{C}$  varsayarak bu regresyonu durdurur: Meta-bilinç, ayrı bir “üst seviye varlık” tarafından değil, aynı  $\pi_t$ - $C_t$  dinamiğinin içsel temsiller üzerinde yeniden uygulanmasıyla ortaya çıkar.

### 8.1.5 Dikkat Şeması Teorisi ile Karşılaştırma

**Dikkat ve bilinç ayrımı.** Dikkat Şeması Teorisi (Attention Schema Theory, AST) [7], bilinci dikkat mekanizmalarının basitleştirilmiş bir iç modelinin yansıması olarak yorumlar. FBÇT ise daha nötr bir ontolojik pozisyon alır:

- Dikkat,  $S_t$ 'nin hangi bileşenlerinin aday üretim çekirdeği  $G$ 'ye ve değer fonksiyonu  $U$ 'ya gireceğini belirleyen bir seçim/ölçekleme mekanizması olarak modellenenir.
- Bilinç ise, yalnızca dikkat süreçlerinin değil, aynı zamanda bellek ( $M_t$ ), değer vektörü ( $W_t$ ) ve çöküş mekanizmasının birlikte ürettiği bir olgudur.

Bu açıdan, AST'nin "dikkat şeması" FBÇT'de  $M_t$  içinde kodlanan özel bir alt-bileşen olarak görülebilir; ancak bilincin tamamını açıklamak için yeterli değildir.

### 8.1.6 Kuantum Bilinç Teorileri ile Karşılaştırma

**Temel uyumsuzluk.** Kuantum bilinç teorileri [11, 12], bilincin özünü kuantum süperpozisyon ve çöküş süreçlerine bağlama eğilimindedir. FBÇT ise şu varsayımları benimser:

- Nöral düzeyde kuantum etkiler tamamen yok sayılmaz, ancak bilinç için *gerekli ve yeterli* oldukları iddia edilmez.
- FBÇT'nin çöküşü tamamen klasik, olasılıksal bir seçim işlemidir; kuantum dalga fonksiyonunun fiziksel çöküşüyle özdeş değildir.
- Teorinin tüm bileşenleri, hesaplanabilir fonksiyonlar ve ölçüler üzerinden tanımlanmıştır; kuantum hesaplama varsayımı yapılmaz.

**Çöküş metaforu.** "Çöküş" terimi benzetmesel olarak kuantum mekaniğinden ödünç alınsa da, FBÇT'de

$$C_t \sim \pi_t(dx) \quad (137)$$

klasik bir olasılık ölçüsü üzerinden örneklemedir. Benzerlik, çoklu olası durumdan tek gerçekleşen duruma geçiş yapısındadır; fiziksel mekanizma ise tamamen farklıdır.

## 8.2 FBÇT'nin Güçlü Yanları

### 8.2.1 Matematiksel kesinlik ve hesaplanabilirlik

FBÇT, bilinci sezgisel metaforlar yerine açık bir hesaplama şeması ile tanımlar. Temel yapı taşları:

Durum:	$(S_t, M_t, W_t, X_t)$	(138)
Aday çekirdeği:	$\mu_t(dx) = G(S_t, M_t, W_t)(dx)$	
Değer fonksiyonu:	$U : X \times \mathcal{S} \times \mathcal{M} \rightarrow \mathbb{R}^k,$ $f(x, S_t, M_t, W_t) = \langle W_t, U(x, S_t, M_t) \rangle$	
Politika:	$\pi_t(dx) \propto \exp(\beta_t f(x, S_t, M_t, W_t)) \mu_t(dx)$	
Çöküş:	$C_t \sim \pi_t(dx)$	
Geri bildirim:	$(M_{t+1}, W_{t+1}) = \mathcal{F}(M_t, W_t, S_t, C_t)$	

Bu yapı üç kritik avantaj sağlar:

1. **Simüle edilebilirlik:** Teori doğrudan algoritmik olarak uygulanabilir (bkz. Algoritma 1).
2. **Tahmin üretme:** Çöküş olasılıkları, bilinç düzeyi metriği  $L_t$  ve parametre duyarlılığı nicel olarak hesaplanabilir.
3. **Model fitting:** Parametreler  $(\beta_t, \eta_M, \alpha_W, \dots)$ , deneysel veriye göre optimize edilebilir.

### 8.2.2 Çok-seviyeli açıklama

FBÇT, bilinç fenomenini Marr [28] tarzı çok-seviyeli bir çerçevede ele alır:

- **Hesaplama seviyesi:** Sistem hangi tür bir fonksiyonu hesaplıyor? (Adaylar üzerinden değer ağırlıklı seçim ve geri bildirim)
- **Algoritma seviyesi:** Bu fonksiyon nasıl gerçekleştiriliyor? (Softmax benzeri politika, aday üretim mekanizmaları, çok ölçekli öğrenme)
- **Uygulama seviyesi:** Bu algoritmalar hangi nöral devreler ve nöromodülatör sistemler üzerinden uygulanıyor? (hipokampus, PFC, amigdala, vb.)

### 8.2.3 Klasik felsefi problemlere sistematik yanıtlar

FBÇT, üç klasik problemi tek bir çerçevede ele alır:

- **Homunculus problemi:** Seçimi yapan ayrı bir “iç gözlemci” yoktur; seçim,  $\pi_t$  ve  $C_t$  ile tanımlanan emergent bir süreçtir.
- **Qualia problemi:** Fenomenolojik içerik,  $(X_t, \mathcal{S})$  üzerindeki eşdeğerlik sınıfları ve  $\Phi$  koordinat fonksiyonu ile temsil edilir; bu yapı davranışsal ve rapora dayalı testlere açılabilir.
- **Özgür irade:** Karar süreçleri deterministik veya olasılıksal olabilir, ancak  $M_t$  ve  $W_t$ ’nin uzun vadeli şekillenmesi, “kendilik sahipliği” (ownership) üzerinden compatibilist bir özgür irade kavrayışı sunar (bkz. Bölüm 5).

### 8.2.4 Geniş uygulama yelpazesi

Teori, yalnızca felsefi bir model değil, pratik alanlara da doğrudan uygulanabilir:

1. **Nörobilim:** Belirli görevlerde  $L_t$ ’nin değişimi, bilinç düzeyiyle ilişkili nöral imzalarla karşılaştırılabilir (Bölüm 6).
2. **Klinik psikiyatri:** Depresyon, anksiyete, travma sonrası stres bozukluğu gibi durumlar,  $M_t$  ve  $W_t$  dinamiklerindeki patolojik sabit noktalar olarak modellenenebilir.
3. **Yapay zeka ve etik:** FBÇT parametrelerini karşılayan yapay sistemler için bilinç ve ahlaki statü tartışmaları formal bir zemine taşınabilir.

## 8.3 Sınırlamalar ve Açık Sorular

### 8.3.1 $\Phi$ fonksiyonu ve hard problem

FBÇT’nin zayıf noktalarından biri, fenomenolojik izdüşüm fonksiyonu  $\Phi$ ’nin henüz tam olarak karakterize edilmemiş olmasıdır. Teoride:

$$(x_1, S_1) \sim (x_2, S_2) \iff \forall g \in \mathcal{G} : g(x_1, S_1) = g(x_2, S_2), \quad (139)$$

ve

$$Q = (X \times \mathcal{S}) / \sim, \quad \Phi : X \times \mathcal{S} \rightarrow \mathbb{R}^m \quad (140)$$

şeklinde bir yapı önerilmiştir. Ancak:

- $\Phi$ ’nin bileşenleri deneysel olarak nasıl tanımlanacaktır?
- Farklı bireyler için  $\Phi$ ’nin biçimi aynı mıdır?
- $\Phi$ ’nin belirli koordinat boyutları, öznel raporlarla nasıl ilişkilendirilecektir?

Bu sorular, hard problem’in ontolojik kısmını çözme de, mekanistik kısmı için açık bir araştırma programı oluşturur (bkz. Bölüm 3.11 ve 6).

### 8.3.2 Parametre tayini ve bireysel farklılıklar

FBÇT çok sayıda parametre içerir (öğrenme hızları,  $\beta_t$ ,  $W_t$  bileşenleri, zaman ölçekleri, vb.). Bunların:

- Türler arası,
- Bireyler arası,
- Gelişimsel dönemler arası,

nasıl değiştiği şu anda büyük ölçüde belirsizdir. Teori, bu parametrelerin deneysel fitting veya bayesyen çıkarım yoluyla tahmin edilmesine olanak tanır; ancak bu program henüz başlangıç aşamasındadır.

### 8.3.3 Hayvan ve gelişimsel bilinç

Model, insan yetişkin bilinci göz önüne alınarak inşa edilmiştir. Hayvanlarda ve gelişimsel aşamalarda  $(S_t, M_t, W_t, X_t)$ 'nin boyutları, yapısı ve dinamikleri farklılık gösterecektir. FBÇT'nin esnekliği, bu farklılıkları teorik olarak kapsamasına izin verse de, somut parametreleşme ve prediktif güç, tür ve yaş grubuna göre yeniden kalibrasyon gerektirir.

### 8.3.4 Yapay bilinç kriteri

Teoriye göre, yeterince zengin bir  $(S_t, M_t, W_t, X_t)$  uzayı ve FBÇT çöküş mekanizmasını implement eden bir yapay sistem, fonksiyonel olarak bilinçli kabul edilmelidir. Ancak şu sorular açık kalır:

- Substrat bağımlılığı var mı? Karbon temelli nöral dokuya özgü özellikler gerekli midir?
- İzomorfik iki farklı implementasyon (biyolojik ve yapay) gerçekten aynı qualia uzayına mı sahiptir?
- Bu tür sistemlerin etik ve hukuki statüsü nasıl belirlenmelidir?

Bu sorular, FBÇT'nin ötesinde, daha geniş bir felsefi ve etik tartışmanın parçasıdır.

## 8.4 Gelecek Araştırma Yönleri

### 8.4.1 Deneysel doğrulama

Bölüm 6'te sunulan nörobiyolojik ve davranışsal tahminler, FBÇT'nin sınanabilirliğinin temelini oluşturur. Özellikle:

- Çöküş anlarında belirli frontoparietal imzaların ve osilatuar desenlerin ortaya çıkması,
- $M_t$  ve  $W_t$  manipülasyonlarının (farmakolojik, davranışsal, nöromodülatif) bilinçli rapor ve seçim olasılıklarını öngörülebilir biçimde değiştirmesi,
- $L_t$  bilinç düzeyi metriğinin, uyanıklık, uyku, anestezi ve meditasyon gibi durumlar arasında sistematik farklar göstermesi,

yüksek öncelikli testlerdir.

### 8.4.2 Hesaplamalı FBÇT ajanı

FBÇT dinamiğini uçtan uca implement eden büyük ölçekli bir hesaplamalı ajan, teoremin hem iç tutarlılığını hem de üretebildiği davranış repertuarını test etmek için kritik önemdedir. Bu ajan, zengin bir çevreyle etkileşime girerek:

- İçgözlem benzeri raporlar,
- Esnek görev uyumu,
- Uzun vadeli değer yeniden kalibrasyonu,

gibi fenomenleri sergileyebilirse, FBÇT'nin yapay sistemler için de geçerli bir açıklama sunduğu iddiası güçlenecektir.

### 8.4.3 Klinik ve terapötik uygulamalar

FBÇT parametreleri, klinik bozuklukları anlamaya yönelik yeni bir dil sunar. Örneğin:

- Depresyon: Belirli değer bileşenlerinin ( $W_t$ ) patolojik ağırlanması,
- PTSD:  $M_t$  içinde travmatik epizotların aşırı güçlenmiş ve kolay tetiklenen attractor havzaları,
- Anksiyete: Tehdit odaklı  $U_i$  bileşenlerinin aşırı baskın olması,

olarak modellenebilir. Bu yaklaşım, hem psikoterapi protokollerinin hem de nöromodülasyon yöntemlerinin (TMS, DBS, vb.) yeniden tasarlanmasına katkı sağlayabilir.

### 8.4.4 Felsefi derinleştirme ve $\Phi$ 'nin açılması

Son olarak, qualia uzayı  $Q$  ve koordinat fonksiyonu  $\Phi$  üzerine daha derin teorik çalışma gereklidir. Burada iki ana eksen öne çıkar:

- **Nörofenomenoloji:** Öznel raporlar ile nöral desenler arasındaki eşlemenin, FBÇT yapısına uygun biçimde istatistiksel olarak modellenmesi.
- **Yapısal kısıtlar:**  $\Phi$  için hangi matematiksel özelliklerin (süreklilik, yerellik, invariantslar) gerekli ve yeterli olduğu sorusu.

## 8.5 Sonuç: FBÇT'nin Yeri

Fonksiyonel Bilinç Çöküşü Teorisi, bilinci:

- Zengin durum uzayları ( $S_t, M_t, W_t, X_t$ ) üzerinde işleyen hesaplanabilir bir seçim süreci,
- Dinamik, çok ölçekli bir geri besleme sistemi,
- Fenomenolojik içeriği eşdeğerlik sınıfları ve koordinat fonksiyonlarıyla temsil edilen bir yapı,

olarak kavramsallaştırır. Teorinin iddiası, bu çerçevenin:

1. Mevcut deneysel verilerle uyumlu,
2. Gelecekteki çalışmalar için zengin ve somut tahminler üreten,
3. Klasik felsefi problemleri (homunculus, qualia, özgür irade) indirgemeci olmayan ama mekanistik bir bakışla yeniden formüle eden,

bir bilinç bilimi altyapısı sunduğudur. FBÇT tamamlanmış bir teori olmaktan uzaktır; ancak sunduğu matematiksel iskelet, hem bilişsel nörobilim hem de yapay zeka araştırmaları için üretken bir başlangıç noktası oluşturmaktadır.

## 9 Kapanış

Fonksiyonel Bilinç Çöküşü Teorisi (FBÇT), bilinci soyut bir gizem alanından çıkararak *hesaplanabilir*, *formel* ve *test edilebilir* bir süreç olarak yeniden çerçevelendirmektedir. Bu çalışma, bilinçli içeriğin duyusal durum ( $S$ ), bellek yapıları ( $M$ ) ve değer önceliklerinin ( $W$ ) etkileşimiyle belirlenen bir seçim dinamiğinden doğduğunu göstermiş; çöküş mekanizması  $\mathcal{C}(S, M, W)$  aracılığıyla hem anlık deneyimi hem de uzun vadeli kimliği açıklayan bütünsel bir model sunmuştur.

FBÇT'nin üç temel katkısı öne çıkmaktadır:

1. **Matematiksel olarak kesin bir çerçeve.** Aday durum uzayları, değer fonksiyonları, politika dağılımı, çöküş operatörü ve geri-besleme dinamikleri formel olarak tanımlanmıştır. Bu yapı, bilinci simüle edilebilir ve deneysel veriye fit edilebilir bir olgu olarak ele alma olanağı sağlar.
2. **Klasik felsefi problemlerin yeniden formüle edilmesi.** Homunculus problemi, hard problem ve özgür irade tartışmaları; fonksiyonel durum uzayları, eşdeğerlik sınıfları ve projeksiyon fonksiyonu  $\Phi$  üzerinden mekanik terimlerle açıklanmıştır. Böylece fenomenolojik yapı metafizik bir muamma olmaktan çıkarak ölçülebilir ve modellenebilir hale gelir.

**3. Disiplinler arası uygulanabilirlik.** FBÇT, nörogörüntüleme, klinik psikiyatri, hesaplamalı modelleme ve yapay zeka alanlarında test edilebilir öngörüler sunmaktadır. Çöküş anlarının nöral imzaları,  $M$  ve  $W$  manipülasyonlarının davranış üzerindeki etkileri ve bilinç düzeyi metriği  $L_t$  için önerilen deneyler, teoriyi yanlışlanabilir bir bilimsel çerçeve haline getirir.

FBÇT tamamlanmış bir teori iddiası taşımamaktadır. Özellikle fenomenolojik projeksiyon fonksiyonu  $\Phi$ 'nin tam formu, bireysel farklılıkların modellenmesi ve türler arası bilinç karşılaştırmaları açık araştırma soruları olarak kalmaktadır. Bununla birlikte, teori bu sorular için net ve çalışılabilir bir araştırma programı sunar.

FBÇT'nin en güçlü yönü, bilinci açıklamak için yeni bir metafizik varsayım önermek yerine, mevcut ampirik yöntemlerin üzerinde çalışabileceği matematiksel bir iskelet sağlamasıdır. Bu iskelet, hem nörobilim hem yapay zeka hem de felsefi analiz için ortak bir dil oluşturur.

Sonuç olarak, FBÇT'nin başarı ölçütü retorik değil, *deneysel veri* olacaktır. Öngörülerini desteklediği ölçüde, bilinç araştırmalarında yeni bir metodolojik çerçeve oluşacaktır; desteklenmediği noktalar ise daha güçlü modeller için yol açacaktır. Her iki durumda da FBÇT'nin katkısı açıktır: bilinci soyut tartışmaların ötesine taşıyarak, bilimsel olarak modellenenebilir bir fenomen haline getirmek.

*Bilinç, yalnızca filozofların değil; matematikçilerin, nörobilimcilerin, bilişimcilerin ve klinisyenlerin birlikte çalışabileceği ortak bir araştırma alanıdır. FBÇT, bu ortak dilin temelini oluşturmayı amaçlamaktadır.*

## Ekler

### A Çöküş Operatörünün Ölçü-Teorik Tanımı

#### A.0.1 Çöküş Operatörünün Matematiksel Tanımı

Bu bölümde çöküş operatörü  $C$ 'nin yalnızca sezgisel değil, ölçü-teorik anlamda da tam bir tanımını veriyoruz. Amaç,  $C$ 'yi “bir seçim kuralı” olmaktan çıkarıp, FBÇT içinde iyi tanımlanmış bir stokastik operatör (Markov çekirdeği) haline getirmektir.

**Aday uzay ve temel ölçü.** Her zaman adımında jeneratif operatör  $G$ , duyuşal, bellek ve öncelik durumlarına bağlı bir aday kümesi üretir:

$$X_t = \{x_t^{(1)}, \dots, x_t^{(K)}\} \subset \mathcal{X}.$$

Basit durumda  $\mathcal{X}$  sonlu veya sayılabilir bir kümedir; sürekli uzay genellemesi aşağıda verilecektir.  $X_t$  üzerinde sayma ölçüsünü  $\mu_t$  ile gösterelim:

$$\mu_t(A) = |\{x \in X_t : x \in A\}|.$$

**Skor ve enerji fonksiyonu.** Bölüm 3.5'te tanımlandığı üzere, her  $x \in X_t$  için zaman-indeksli skor fonksiyonu:

$$f_t(x) \equiv f(x; S_t, M_t, W_t)$$

ve buna karşılık gelen enerji fonksiyonu:

$$E_t(x) := -f_t(x).$$

Enerji, adayın duyuşal uygunluk, bellek uyumu ve motivasyonel değer bileşenlerine göre “ne kadar maliyetli” olduğunu temsil eder.

**Gibbs/Boltzmann dağılımı.** Çöküş dağılımı, enerji fonksiyonu üzerine tanımlı bir Gibbs ölçüsüdür. Sıcaklık parametresi  $\beta_t > 0$  için:

$$\pi_t(x \mid S_t, M_t, W_t) := \frac{\exp(-\beta_t E_t(x))}{\sum_{j=1}^K \exp(-\beta_t E_t(x_t^{(j)}))} = \frac{\exp(\beta_t f_t(x))}{\sum_{j=1}^K \exp(\beta_t f_t(x_t^{(j)}))}.$$

Bu,  $X_t$  üzerinde tanımlı bir olasılık ölçüsüdür:

$$\pi_t(\cdot \mid S_t, M_t, W_t) : \mathcal{P}(X_t) \rightarrow [0, 1], \quad \sum_{x \in X_t} \pi_t(x \mid S_t, M_t, W_t) = 1.$$

**Çöküş operatörü olarak Markov çekirdeği.** Çöküş operatörü  $C$ ,  $(S_t, M_t, W_t)$  üçlüsünü girdiler ve  $X_t$  üzerindeki olasılık dağılımını çıktı olarak alan bir Markov çekirdeği olarak tanımlanır:

$$C_t : (S_t, M_t, W_t) \mapsto \Pi_t,$$

burada

$$\Pi_t(A; S_t, M_t, W_t) := \sum_{x \in A} \pi_t(x \mid S_t, M_t, W_t), \quad A \subseteq X_t.$$

Böylece  $C_t$ , her bağlamsal durum için aday kümesi üzerinde iyi tanımlanmış bir olasılık ölçüsü üretir.

Bilinçli içerik  $C_t$  bir rassal değişkendir:

$$C_t \sim \pi_t(\cdot \mid S_t, M_t, W_t),$$

yani

$$\mathbb{P}[C_t = x \mid S_t, M_t, W_t] = \pi_t(x \mid S_t, M_t, W_t).$$

**Süreklili aday uzayı genellemesi.** Aday uzayın sürekli olduğu durumda,  $(\mathcal{X}, \mathcal{B})$  ölçülebilir bir uzay ve  $\mu$  uygun bir temel ölçü (örn. Lebesgue ölçüsü veya önceden tanımlı bir referans dağılım) olsun. Enerji fonksiyonu  $E_t : \mathcal{X} \rightarrow \mathbb{R}$  ölçülebilir ve aşağıdaki Gibbs yoğunluğunu tanımlar:

$$p_t(x \mid S_t, M_t, W_t) = \frac{1}{Z_t} \exp(-\beta_t E_t(x)),$$

burada normalizasyon sabiti:

$$Z_t = \int_{\mathcal{X}} \exp(-\beta_t E_t(x)) \mu(dx).$$

Bu durumda çöküş operatörü,  $\mathcal{B}$  üzerinde tanımlı bir Markov çekirdeği olarak yazılabilir:

$$\Pi_t(A; S_t, M_t, W_t) := \int_A p_t(x \mid S_t, M_t, W_t) \mu(dx), \quad A \in \mathcal{B}.$$

**Sınır rejimleri.** Aynı formalizm, farklı bilişsel rejimleri tek parametre ile kapsar:

$$\beta_t \rightarrow 0 \Rightarrow \pi_t(x) \approx \text{uniform}(X_t),$$

$$\beta_t \rightarrow \infty \Rightarrow \pi_t(x) \rightarrow \mathbb{I}\{x \in \arg \max f_t\},$$

yani çok düşük sıcaklıkta çöküş deterministik argmax'a yaklaşır, çok yüksek sıcaklıkta ise neredeyse rastgele örnekleme rejimine geçilir.

**Zaman dinamiği ile entegrasyon.** FBÇT'de çöküş, yalnızca o anki bilinç içeriğini üretmekle kalmaz, aynı zamanda bellek ve öncelik sistemlerinin gelecekteki durumlarını belirleyen güncelleme operatörüne girdi sağlar:

$$(M_{t+1}, W_{t+1}) = F(M_t, W_t, S_t, C_t, R_t),$$

burada  $R_t$  uygun bir ödül/ceza sinyalıdır. Böylece  $C_t$ , hem  $(S_t, M_t, W_t)$ 'den türeyen hem de  $(M_{t+1}, W_{t+1})$ 'yi şekillendiren merkezi bir stokastik düğüm haline gelir.

## A Kaynakça

## References

- [1] David J Chalmers. Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3): 200–219, 1995.
- [2] Thomas Nagel. What is it like to be a bat? *The philosophical review*, 83(4):435–450, 1974.
- [3] Daniel C Dennett. *Consciousness explained*. Little, Brown and Co, Boston, 1991.
- [4] Bernard J Baars. A cognitive theory of consciousness. 1988.

- [5] Stanislas Dehaene and Lionel Naccache. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1-2):1–37, 2001.
- [6] David M Rosenthal. Consciousness and mind. 2005.
- [7] Michael SA Graziano. Consciousness and the social brain. 2013.
- [8] Ned Block. On a confusion about a function of consciousness. *Behavioral and brain sciences*, 18(2): 227–247, 1995.
- [9] Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience*, 5(1):1–22, 2004.
- [10] Giulio Tononi, Melanie Boly, Marcello Massimini, and Christof Koch. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461, 2016.
- [11] Roger Penrose. *The emperor’s new mind: Concerning computers, minds and the laws of physics*. Oxford University Press, Oxford, UK, 1989.
- [12] Stuart Hameroff and Roger Penrose. Orchestrated reduction of quantum coherence in brain micro-tubules: A model for consciousness. *Mathematics and computers in simulation*, 40(3-4):453–480, 1996.
- [13] Stanislas Dehaene and Jean-Pierre Changeux. Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227, 2011.
- [14] Michael SA Graziano and Taylor W Webb. The attention schema theory: a foundation for engineering artificial consciousness. *Frontiers in Robotics and AI*, 2:60, 2015.
- [15] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2): 127–138, 2010.
- [16] Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204, 2013.
- [17] Anil K Seth. Being you: A new science of consciousness. 2021.
- [18] Bernard J Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, Cambridge, UK, 1988. İngilizce orijinal baskı; Global Neuronal Workspace yaklaşımının köken metni.
- [19] Stanislas Dehaene. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking, New York, 2014.
- [20] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008.
- [21] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annual review of neuroscience*, 30:535–574, 2007.
- [22] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6): 1161–1178, 1980. doi: 10.1037/h0077714.
- [23] Francisco J Varela, Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA, 1991. ISBN 9780262720212.
- [24] Francisco J Varela. Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3(4):330–349, 1996.
- [25] Benjamin Libet, Curtis A Gleason, Elwood W Wright, and Dennis K Pearl. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain*, 106(3):623–642, 1983.
- [26] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5):543–545, 2008.



- [27] Daniel C Dennett. *Freedom evolves*. Penguin, New York, 2004.
- [28] David Marr. Vision: A computational investigation into the human representation and processing of visual information. 1982.