תרגול 6
שיפורים לאלפא-ביטא
משחקים הסתברותיים
משחקים עם אינפורמציה חלקית

מבוא לבינה מלאכותית (236501)

מדעי המחשב, טכניון

אביב 2023

היום בתרגול

- חזרה קצרה על אלגוריתם $\boldsymbol{MiniMax}$ ו- $\boldsymbol{\alpha\beta}$
- שיפורי $\boldsymbol{\alpha\beta}$:
  - Anytime, סידור בנים, העמקה סלקטיבית, ספריות פתיחה/סיום, טבלאות מצבים
- משחקים הסתברותיים
  - אלגוריתם Expectimax
- משחקי ידע חלקי
  - אלגוריתם Monte-Carlo
- שאלות סיכום ממבחנים בנושא משחקים

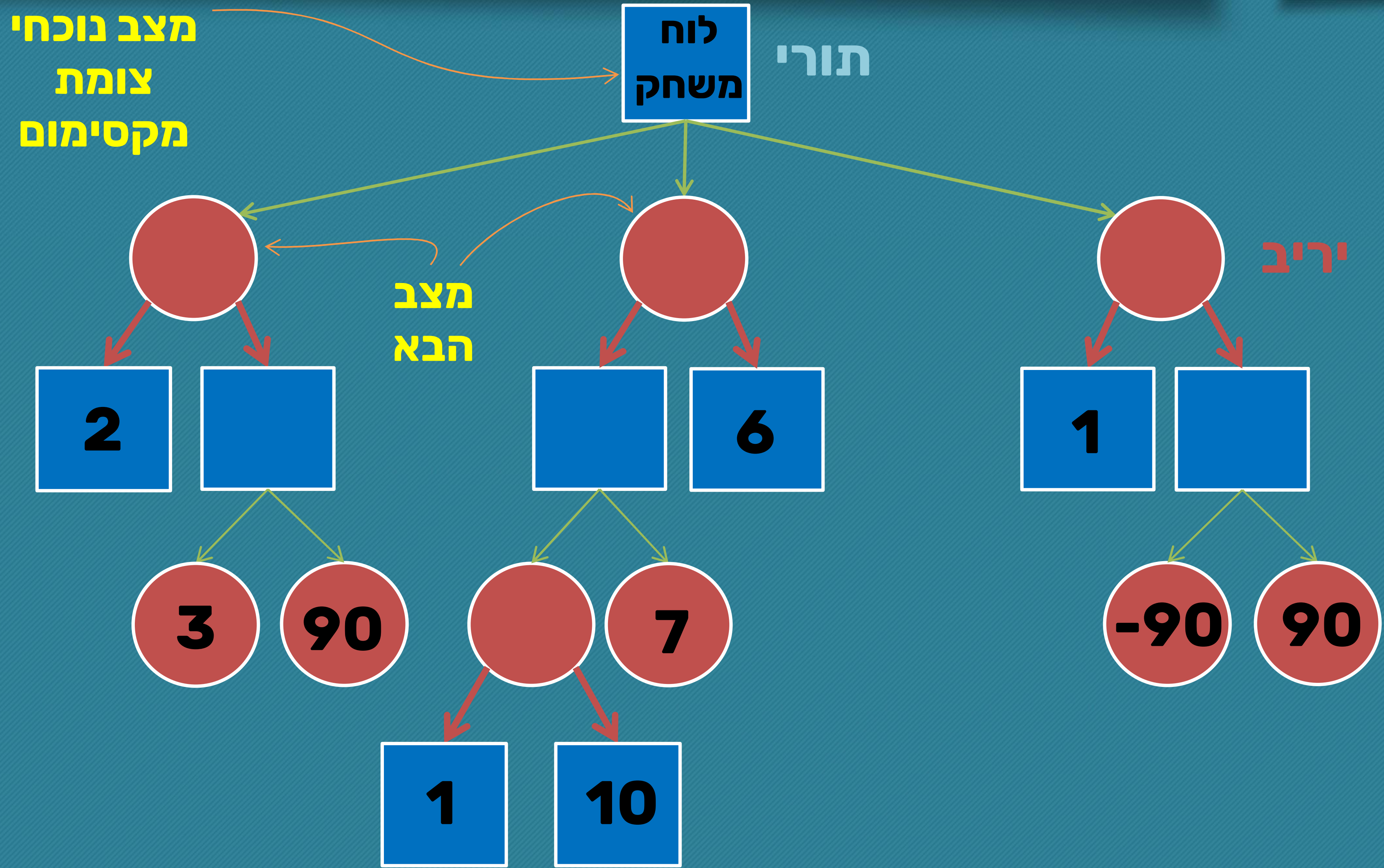
תזכורת: פונקציית Minimax

- **פונקציה** מוגדרת היטב $\mathbb{R} \mapsto [n] \times S : Minimax$
- **מרחיבה** את $\mathbb{R} \mapsto [n] \times S_G : U$ עבור מצבים לא סופיים.
- מניחה **חוסר ידע** מוחלט על **פעולות** היריב.
- מוגדרת רקורסיבית:

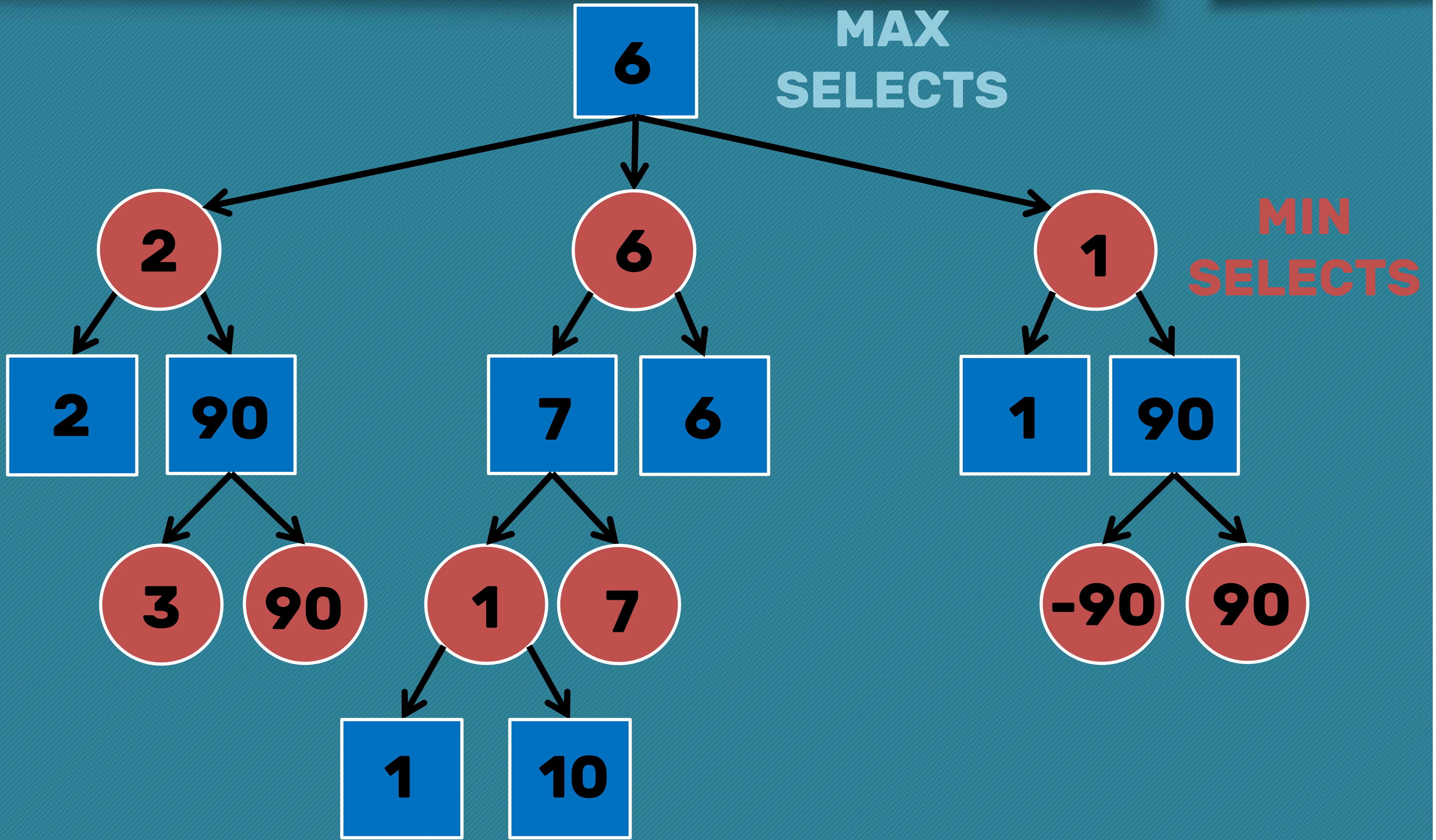
$$Minimax(s,k) \triangleq \begin{cases} U(s,k) & P_G(s) \\ \max_{s' \in Succ(s)} \{Minimax(s',k)\} & Turn(s) = k \\ \min_{s' \in Succ(s)} \{Minimax(s',k)\} & Turn(s) \neq k \end{cases}$$

- משמעות: בהינתן **מצב** s , **שחקן k יכול להגיע** למצב סופי g בעל **תועלת לפחות** $U(g) \geq Minimax(s,k)$ ו**ללא תלות** בבחירות **היריב**.

תזכורת: RB-Minimax



תזכורת: דוגמת הרצת RB-Minimax



תזכורת: גיזום $\boldsymbol{\alpha\beta}$

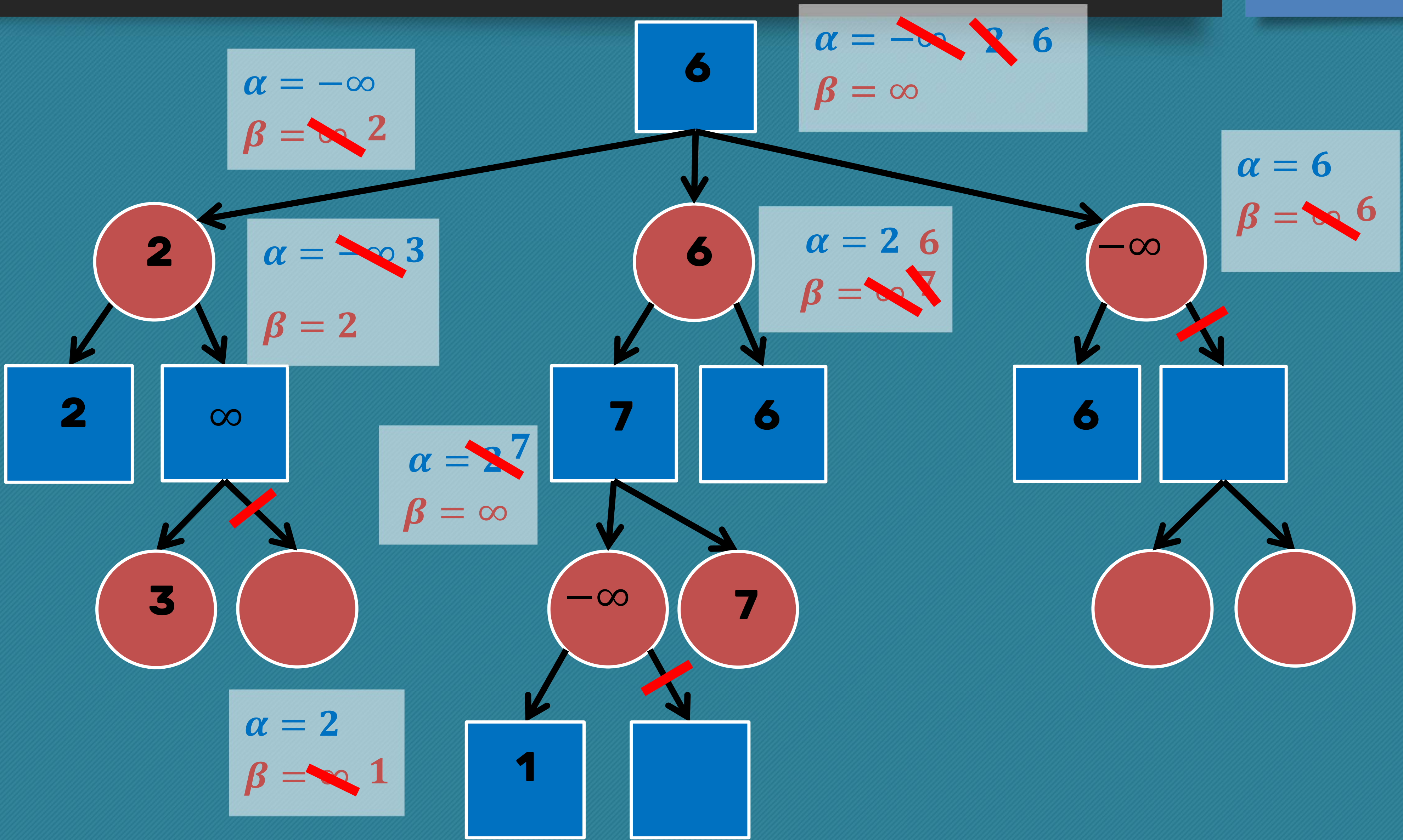
קיצוץ ענפים שפיתוחם בהכרח לא ישנה את ערך המינימקס, מסלול המינימקס ואסטרטגיית המינימקס.

הרעיון: נעביר כלפי מטה בקריאה הרקורסיבית **שני חסמים** $\boldsymbol{\alpha,\beta}$

במהלך הריצה:

- α – **המקסימום** מבין הערכים המובטחים לאבות קדמונים שהם צמתי **max**
  - הערך של האב הקדמון יהיה גדול-שווה α
- β – **המינימום** מבין הערכים המובטחים לאבות קדמונים שהם צמתי **min**
  - הערך של האב הקדמון יהיה קטן-שווה β

תזכורת: דוגמת הרצת $\alpha\beta$



שיפורי $\alpha\beta$

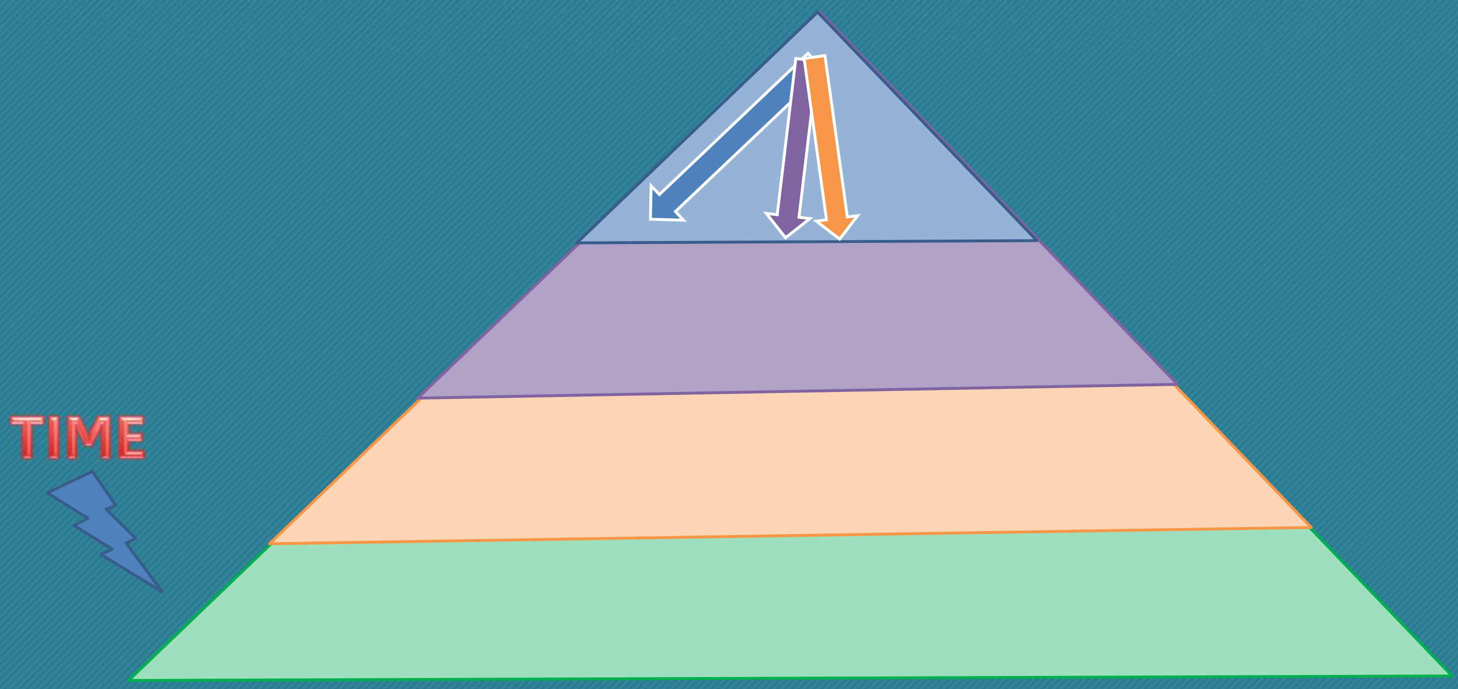
Anytime, סידור בנים, העמקה סלקטיבית, ספריות
פתיחה/סיום, טבלאות מצבים

שיפורי $\alpha\beta$ – anytime

הגדרת בעיה מתוקנת:
במקום "החזר את הצעד הטוב ביותר" –
"החזר את הצעד הטוב ביותר שאתה יכול למצוא תוך k שניות"

- איך נריץ $\alpha\beta$?
  - נבצע העמקה הדרגתית (Iterative deepening)
- כמה איטרציות?
  - נעצור כשיגמר הזמן.
- איזה פתרון נחזיר כשנגמר הזמן?
  - של האיטרציה הקודמת שנסרקה במלואה.

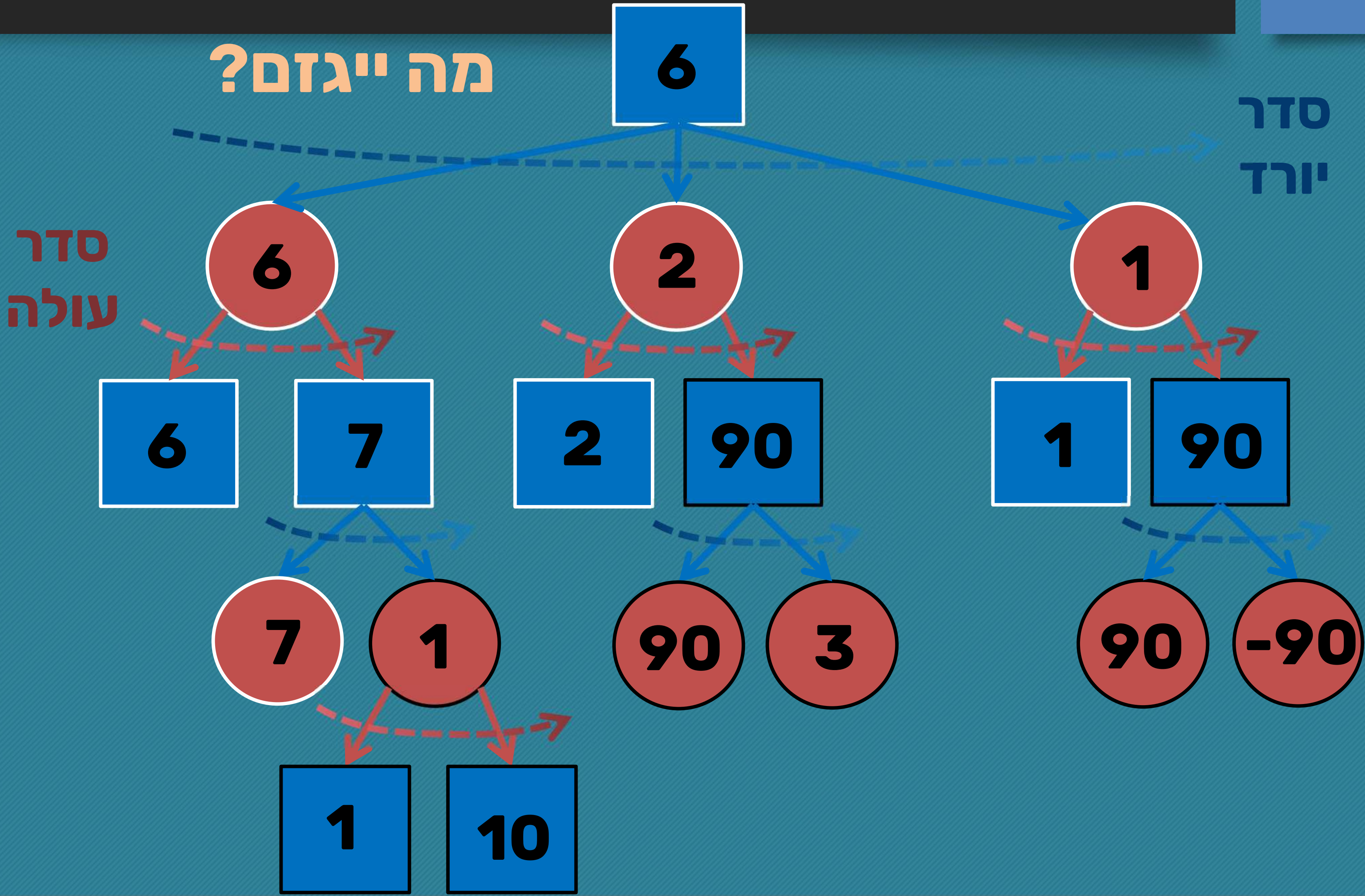
שיפורי $\alpha\beta$ – anytime – הדגמה



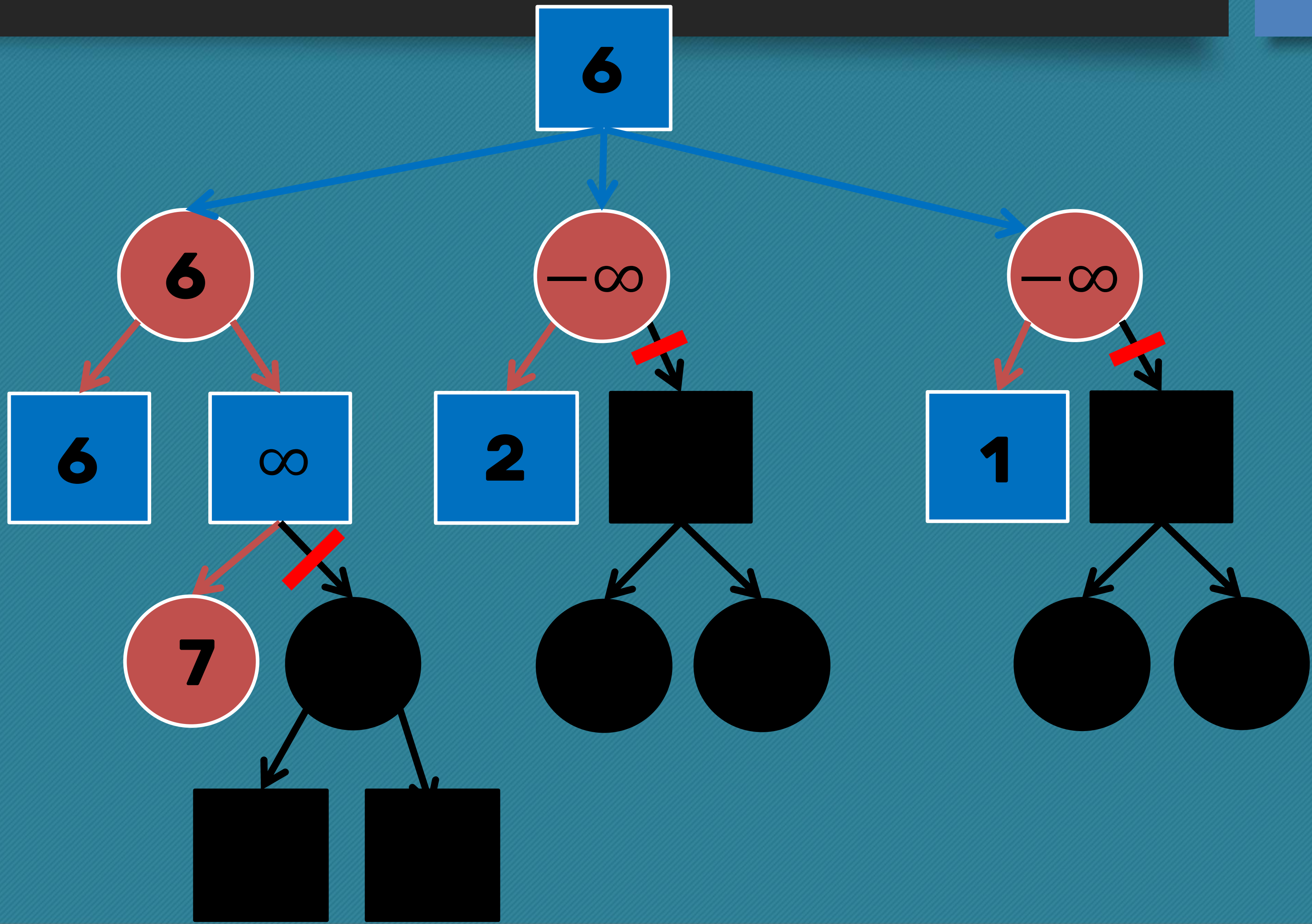
שיפורי $\boldsymbol{\alpha\beta}$ – סידור ילדים

- כדי למקסם את כמות הגיזומים, באיזה **סדר** היינו רוצים **לעבור על הילדים** של צומת?
  - בנים של צמתי **max** בסדר **יורד**
    - כך אם ערך המינימקס של אחד הבנים גדול מ- β נדע זאת לאחר פיתוח הבן הראשון, ונגזום.
    - כך אולי נקבל חסם אלפא גדול עבור המשך שאר הילדים של הצומת.
  - בנים של צמתי **min** בסדר **עולה**

שיפורי $\alpha\beta$ – סידור ילדים – דוגמה



שיפורי $\alpha\beta$ – סידור ילדים – דוגמה



שיפורי $\alpha\beta$ – סידור ילדים

<u>בעיה</u>: רוצים לסדר את הבנים לפי הערך שלהם, כדי **לחסוך** את חישוב הערך שלהם.

אז.. **איך נדע** את ערכי הבנים **מבלי לחשב** אותם?

- נשערך אותם בעזרת היוריסטיקה!

כלומר בכל מצב במהלך הריצה, **נשערך** את כל המצבים העוקבים שלו, נמיין אותם בסדר הרצוי ונפתח אותם לפי הסדר.

  - עלול להיות יקר.. צריך היוריסטיקה קלה לחישוב.

- אם אנחנו בגרסת anytime: נמיין את הבנים לפי ערכם באיטרציה הקודמת.

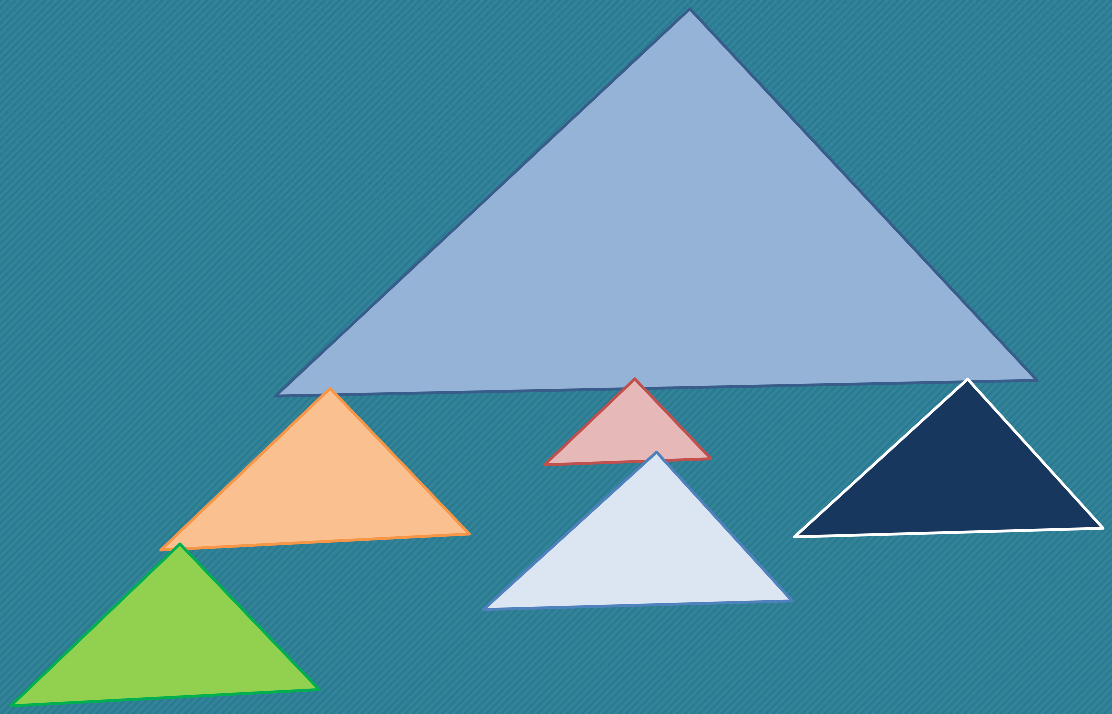
שיפורי $\alpha\beta$ – העמקה סלקטיבית

- הגענו למצב שברור שהוא **קרוב לסיום** המשחק.
- אבל.. הגענו **לגבול ההעמקה** (עבור איטרציה זו).
- **אילו רק** היינו יכולים לעשות **עוד צעד אחד**...
- היינו נותנים הערכה מדויקת.

שיפורי $\alpha\beta$ – העמקה סלקטיבית

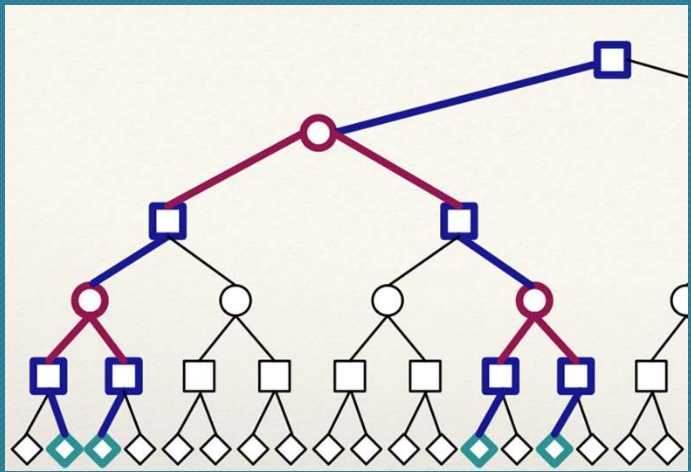
- רעיון: בתנאי מסוים, נאפשר להעמיק עוד קצת.
- באיזה תנאי?
  - מצב "לא שקט"
  - כאשר הצעד האחרון שבוצע "מהפכני"
  - רעיון למדד כזה?
  - למשל: **הפרש ערך היוריסטי** גדול בין מצבים עוקבים
- כמה נמשיך להעמיק?
  - עד לרגיעה, לכל היותר עד עומק D+k

שיפורי $\boldsymbol{\alpha\beta}$ – העמקה סלקטיבית



שיפורי $\alpha\beta$ – ספריות פתיחה / סיום

- נזכור **רצפי משחק/תוכניות פעולה** למצבים נפוצים, ונשתמש בהם אם נגיע למצב שהוא חלק מרצף כזה.

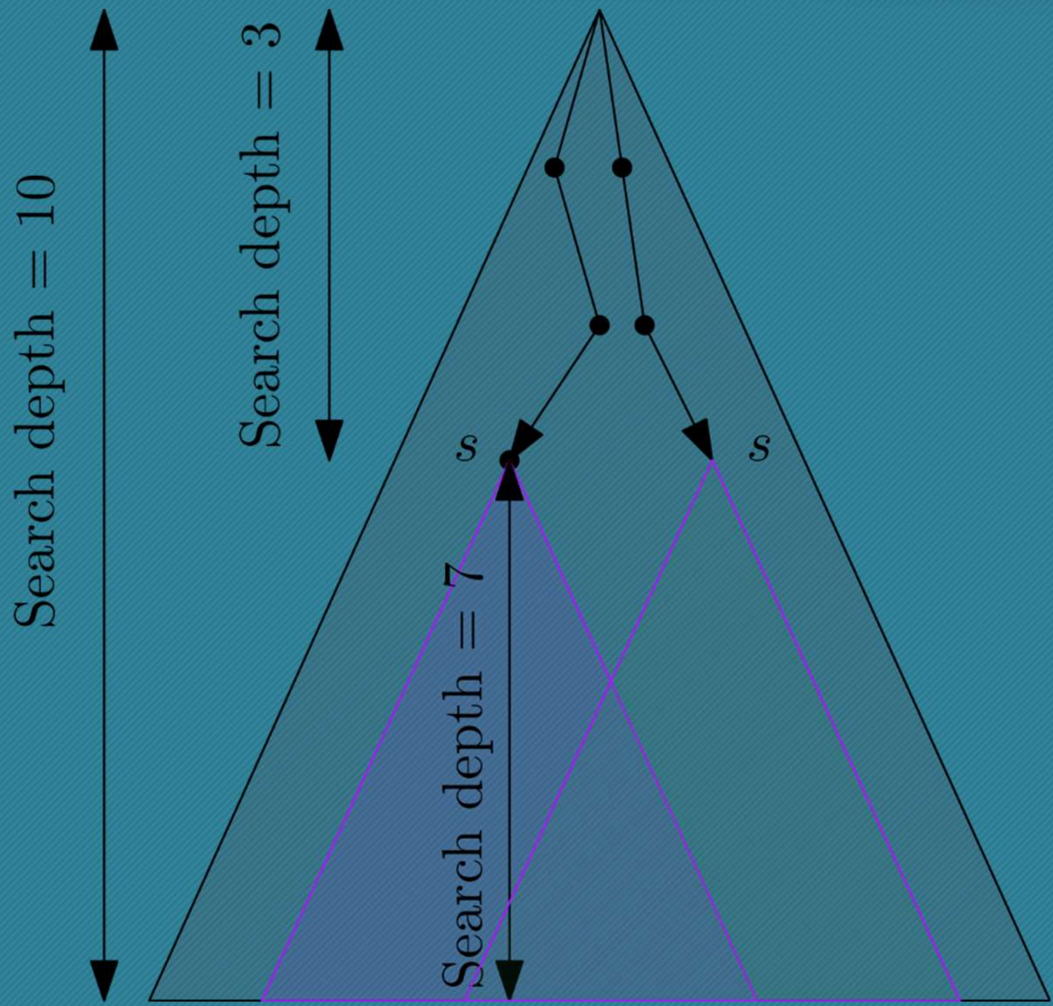


- למה פתיחות (בשחמט)...?
  - מספר מצומצם של פתיחות מקובלות – סיכוי טוב שנשתמש בתכנית הפעולה שחישבנו מראש.
- למה סיומות (בשחמט)...?
  - מספר מועט של כלים – מקדם סיעוף קטן - צריך פחות זיכרון כדי לשמור את תכנית הפעולה.

שיפורי $\alpha\beta$ – טבלאות מצבים

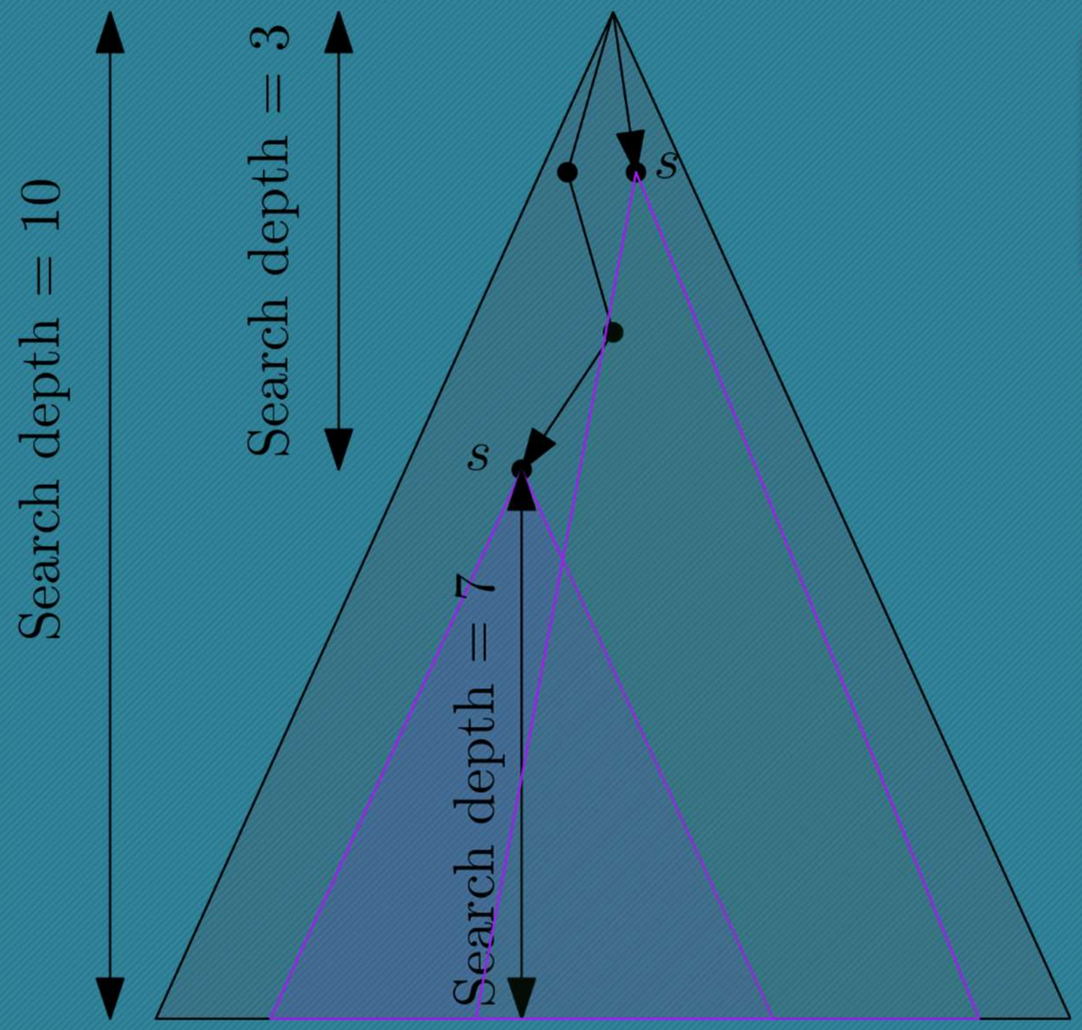
- במהלך החיפוש נשמור את ערכי המינימקס של המצבים בטבלה, עם רשומות מהצורה:
< **מצב**, עומק החיפוש שהערך משקף, **ערך מינימקס** >
- מתי נשתמש בערך השמור בטבלה?
  - אם **מתישהו** (אולי בתור אחר) ניפגש שוב באותו **המצב**
- למה משמש השדה "עומק החיפוש שהערך משקף"?
  - לרוב, נשתמש בערך מהטבלה עבור מצב רק אם עומק החיפוש שהערך משקף => מהעומק שנותר בקריאה הנוכחית למינימקס.
  - אחרת, במינימקס לעומק d האסטרטגיה תיקבע על פי ערכי מינימקס מעומקים קטנים מ- d.
  - **כדי לקיים את משפט ההבטחה של מינימקס**, צריך להשתמש בערך מהטבלה רק אם עומק החיפוש שהערך משקף שווה בדיוק לעומק החיפוש שנותר.

שיפורי $\alpha\beta$ – טבלאות מצבים



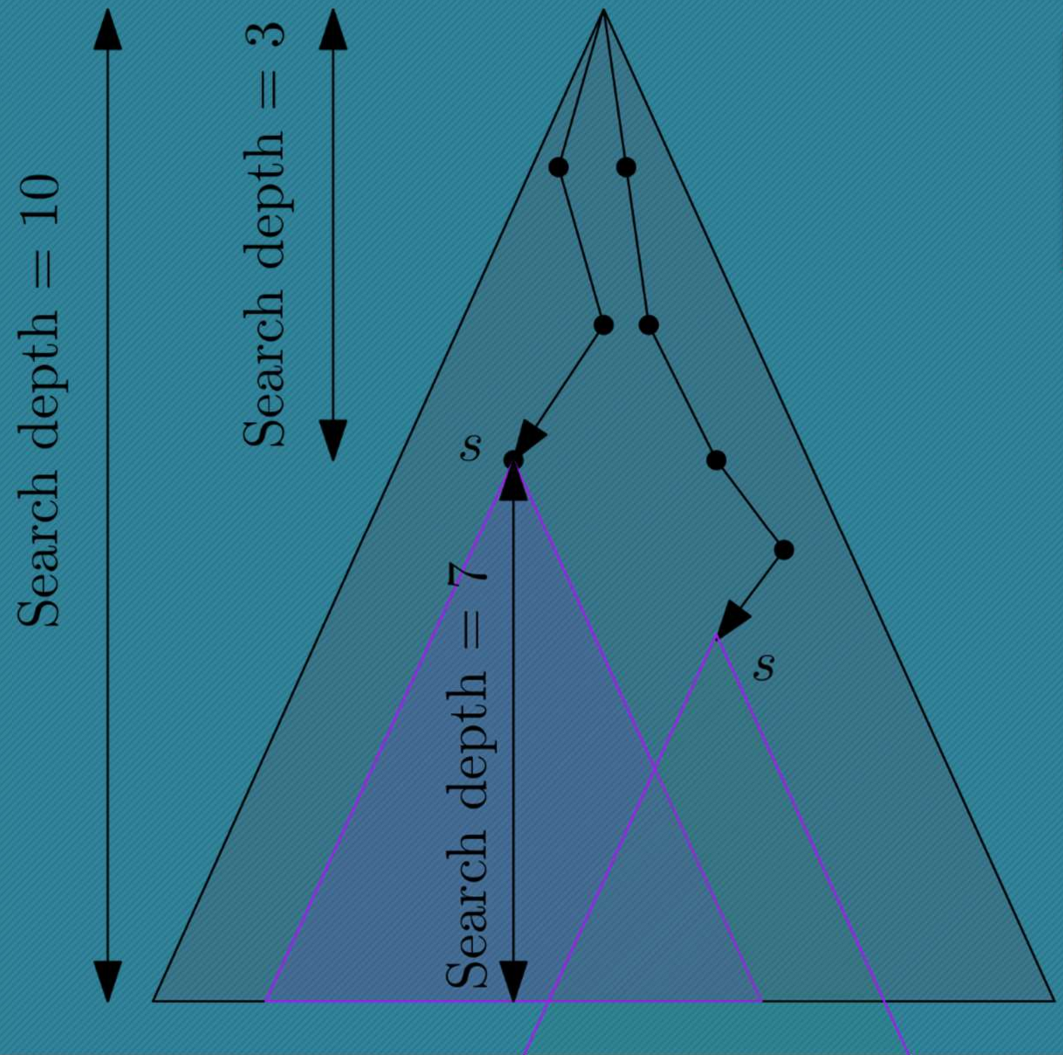
state	depth	value
s	3	75

שיפורי $\alpha\beta$ – טבלאות מצבים



state	depth	value
s	3	75

שיפורי $\alpha\beta$ – טבלאות מצבים



state	depth	value
s	3	75

משחקים **הסתברותיים**

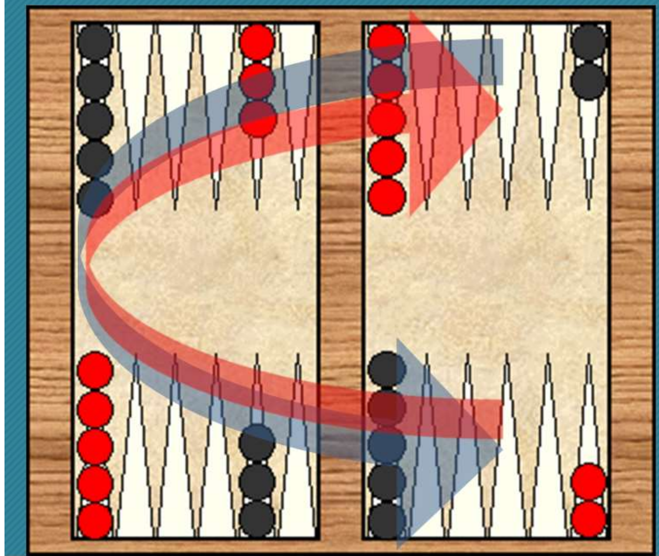
שש-בש – חוקים

<u>מצב התחלתי</u>: קבוע וידוע מראש.

<u>המטרה</u>: "הוצאת" כל הכלים של השחקן מהלוח

<u>האמצעי</u>: הנעת הכלים עם כיוון השעון.

לאחר שכל הכלים הועברו לרביע האחרון של שחקן, הוא מתחיל להוציא אותם.



שש-בש

הציעו היוריסטיקה פשוטה עבור משחק השש-בש

נרצה לבטא את המרחק של כלל חיילי שני השחקנים מסוף הלוח.

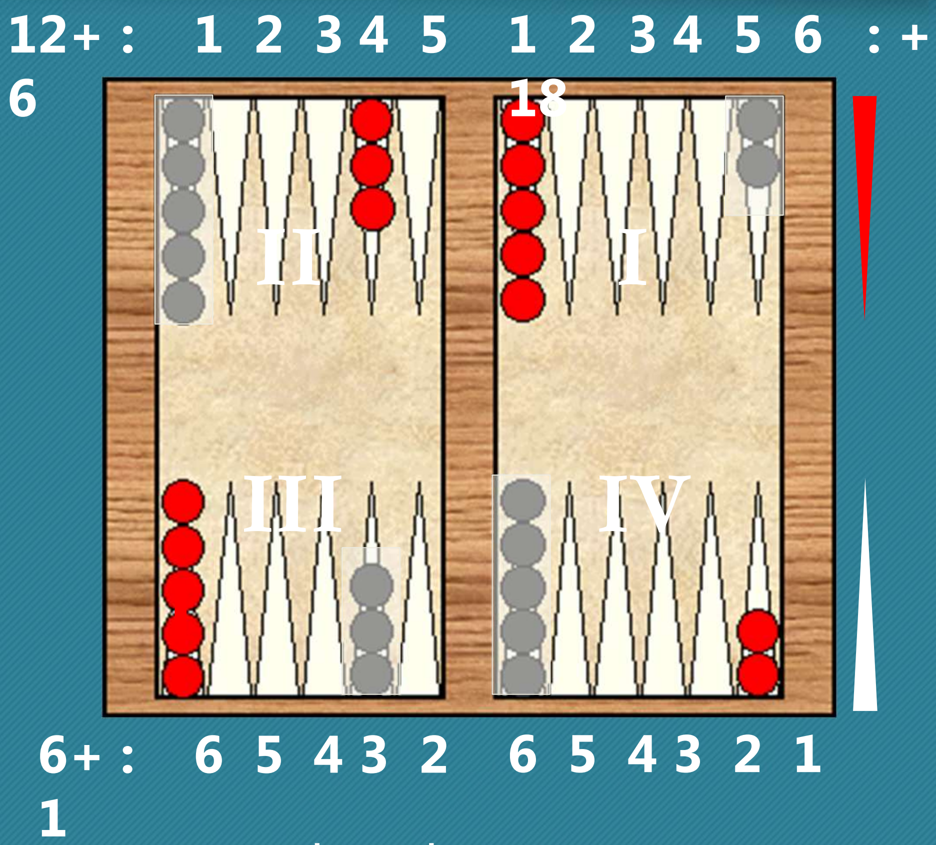
לצורך כך נמספר את צעדי היחידה מכל נקודה בלוח עד ל"יציאה".

דוגמת חישוב

$$5 \cdot 6$$
$$3 \cdot 8$$
$$5 \cdot 13$$
$$2 \cdot 24$$

$$\Sigma_{Black} = 167$$
$$\Sigma_{Red} = \Sigma_{Black}$$

$$h(S_i, B) = \Sigma_R - \Sigma_B = 0$$



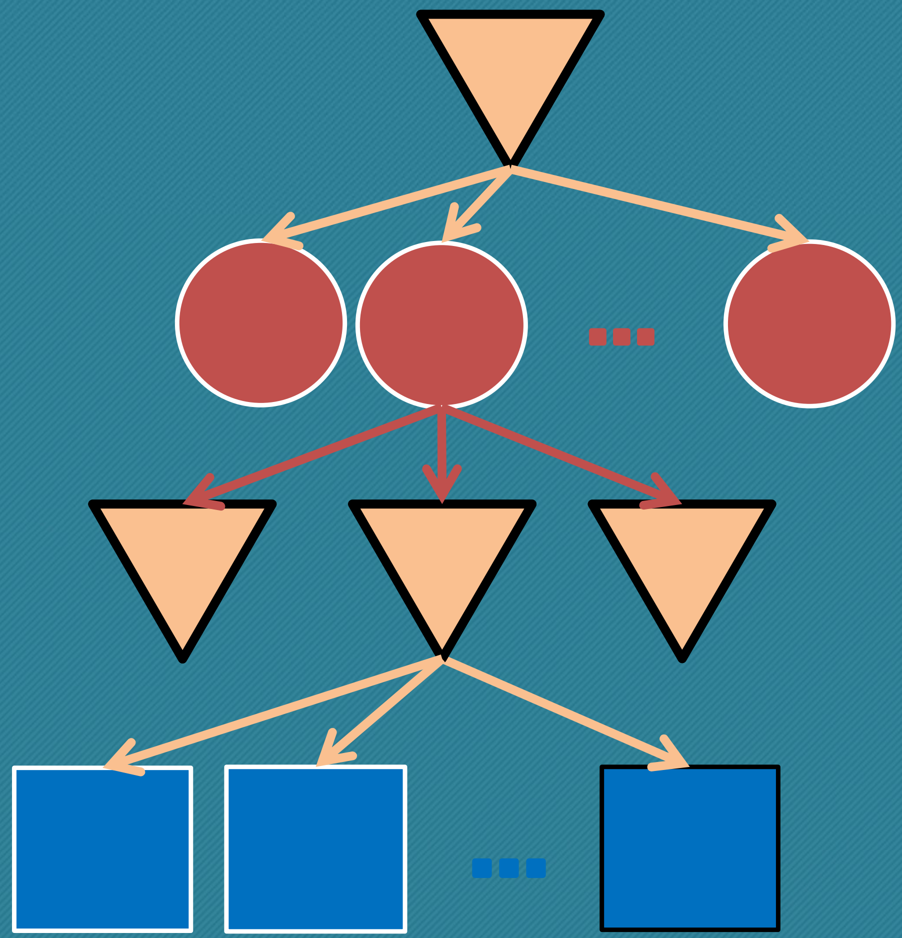
חוץ = 25

נסכם את מרחקי כל חיילי היריב מהיציאה ונחסיר את הסכום המקביל עבור חיילי הסוכן.

שש-בש: איך נייצג את **עץ המשחק**?

מאיפה מגיעות תוצאות ה**קובייה**?

שש-בש – ייצוג הקובייה כסוכן



שש-בש – ייצוג **הקובייה** כ**סוכן**

- <u>מינימקס</u>: האם הגיוני להשתמש בערך **מינימלי** גם על צמתי **קובייה**?
- יש לנו **ידע** על ה**קובייה**:
  - יש להטלות **התפלגות** ידועה.
  - לעומת ה**יריב**, ה**קובייה** לא מנסה לנצח אותנו.
- <u>רעיון</u>: נחשב **תוחלת** על ערכי המינימקס של הבנים, לפי **התפלגות** ההטלות.

function RB-Expectimax(State, Agent, D):

if G(State) **OR** D=0 then **return** h(State, Agent)

if Probablistic(State) **then**

Return $\sum_{<c,p>\in Prob(State)} p \cdot$ RB-Expectimax(c,Agent,D-1)

Turn ← Turn(State)

Children ← Succ(State)

if Turn = Agent then:

CurMax ← $-\infty$

Loop for c in Children

v ← RB-Expectimax(c, Agent, D-1)

CurMax ← Max(v, CurMax)

Return(CurMax)

else: (Turm ≠ Agent)

CurMin ← ∞

Loop for c in Children:

v ← RB-Expectimax(c, Agent, D-1)

CurMin ← Min(v, CurMin)

Return(CurMin)

DIFF (with RB-Minimax)

שאלה ממבחן

חורף 2007-8, מועד א

.2

א. אילו שינויים נדרשים באלגוריתם RB-Expectimax עבור משחקים הסתברותיים כמו שש-בש כדי שמשפט ההבטחה של המינימקס מוגבל המשאבים יהיה נכון עבורו. נמקו.

האלגוריתם מחשב תוחלת בצמתים הסתברותיים, כלומר סכום משוקלל של ערכי RB-Expectimax לפי ההסתברויות השונות.

משפט ההבטחה מבטיח שלאחר D צעדים נגיע למצב בו הערך היוריסטי הוא לפחות הערך המוחזר.

על מנת להבטיח זאת, צריך לבחור בצמתים ההסתברותיים את הערך הנמוך ביותר במקום התוחלת. כלומר, להתייחס לצמתים ההסתברותיים כצמתי מינימום נוספים.

שאלה ממבחן

חורף 2007-8, מועד א

השווה את התנהגותו של האלגוריתם המעודכן לזו של RB-Expectimax המקורי.

שחקן זה ישחק באופן שמרני (פחדני) כיוון שהוא מניח את המקרה הגרוע עבור הגורמים ההסתברותיים וכך ימנע ממהלכים אשר בהסתברות גבוהה יניבו תוצאות טובות רק בגלל החשש מהמקרה הגרוע והפחות סביר. ככל הנראה השחקן ישחק בצורה פחות טובה מהאלגוריתם המקורי.

משחקי אינפורמציה חלקית

אלגוריתם Monte-Carlo

משחקי אינפורמציה חלקית

- **מצב** המשחק ידוע רק **חלקית** ל**שחקן** הנוכחי.
- איך אחשב מה יעשה ה**יריב** אם **לא ידועים** לי את הקלפים שלו?
- רעיון:
  - **נשלים** את המצב החלקי ל**מצב מלא** בכל דרך חוקית
  - **נריץ** אלפא-ביתא על כל **מצב מלא** כזה
  - נעשה **ממוצע** על כל צעד אפשרי
  - נבחר את **הצעד** עם ה**ממוצע המקסימלי**

משחקי אינפורמציה חלקית

- <u>בעיה?</u>
  - יותר מידי דרכים להשלים!
- <u>פתרון:</u>
  - נדגום רק k אפשרויות, במקום לרוץ על כולם.

Monte-Carlo אלגוריתם

function MonteCarlo(PartialState, Agent, D, **K**):

Actions ← Legal actions in PartialState

S ← All states consistent with PartialState

Sample ← Random **K** elements of S

Loop for a in Actions

V(a) ←AVG RB-AlphaBeta(a(s), Agent, D, Alpha=−∞, Beta=∞) **for** s in Sample

Select a with maximal V(a)