systems

MDPI

Article

Systems Thinking with Causal Loop Diagrams in Medical Education: An Exploratory Study

David M. Rubin [1,*], Shamin Achari [1], Xriz L. Richards [1], Adam Pantanowitz [1,2] and Ann George [3,4]

[1] Biomedical Engineering Research Group, School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg, 1 Jan Smuts Avenue, Braamfontein 2001, South Africa
[2] Wits Innovation Centre, University of the Witwatersrand, Johannesburg, 1 Jan Smuts Avenue, Braamfontein 2001, South Africa
[3] Centre for Health Science Education, Phillip Tobias Building, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 29 Princess of Wales Terrace, Parktown 2193, South Africa
[4] Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK
* Correspondence: david.rubin@wits.ac.za

Abstract

Medical literature is replete with diagrammatic representations of systems, yet lacks standardised nomenclature and consistent symbolic conventions. In an introductory system dynamics course for health science students, causal loop diagrams (CLDs) are used to support systems thinking. Notwithstanding recognised limitations, CLDs provide a coherent heuristic for representing multivariate systems with feedback. We studied 55 first-year volunteers enrolled in the course to compare understanding of systems presented as CLDs versus typical journal diagrams. Two endocrine systems were selected from open-access, peer-reviewed literature: calcium homeostasis and glucose homeostasis. Participants were shown either the original journal diagram for one system and a CLD for the other, or vice versa, and answered twelve true/false questions—six per system. A mixed-model, two-way repeated measures ANOVA revealed a significant interaction between Diagram (CLD vs. journal diagram) and System (Calcium vs. Glucose). Post hoc comparisons showed significantly higher performance with CLDs for both Calcium (0.84 vs. 0.38) and Glucose (0.83 vs. 0.63), $p < 0.001$. A Fisher's Exact Test also indicated a higher proportion of questions favouring CLDs. These findings suggest that training in CLDs may enhance understanding of complex systems compared to standard journal diagrams. Further work is needed to address limitations including the small sample size, use of a single cohort, and a restricted set of diagrams.

Keywords: causal loop diagrams; medical illustrations; systems thinking; system dynamics; medical education

Academic Editor: Roger Hadgraft

Received: 27 February 2026
Revised: 23 March 2026
Accepted: 30 March 2026
Published: 1 April 2026

1. Introduction

The study of medicine is, to a large extent, the study of complex systems. The theories and methods of systems science provide a rich and powerful framework for engaging with the dynamic, nonlinear, and interdependent nature of medical practice. However, medical education is not structured like engineering training, and medical students typically do not receive formal instruction in systems theory or systems-based analytical approaches.

A textual description of a system is almost always enriched by the inclusion of diagrammatic representations. Systems diagrams, of which there are many types, ideally capture the essence of a system's structure and function in a compact and economical

way. Historically, illustrations were frequently employed to enhance medical texts [1]. While many of these early diagrams are anatomical, some are synoptic or summary diagrams [1]—a class of illustration that could be described as rudimentary systems diagrams. Similarly, contemporary medical textbooks and journals are filled with diagrams representing a wide range of systems, and attempts have been made to classify them [2]. However, there remains substantial uncertainty as to the characteristics required to optimally convey information in medical illustrations and diagrams [3]. Tippett examines the role of the diagram as a tool to "learn with" rather than to "learn from" [4]—a distinction that captures an important aspect of systems representations in education.

It has been argued that systems modelling is the very essence of science—the basis on which we visualise, organise, simplify, and structure our ideas in order to examine, share, and extend our knowledge [5]. Systems thinking has been described as a unifying paradigm for biological sciences, with implications for teaching as well as research [6], and the role of systems modelling in teaching conceptually difficult physiological concepts has been recognised [7]. The importance of understanding processes in biology is well described [8–10], and diagrammatic representations are a key component in achieving clarity [8]. However, despite attempts to identify suitable symbols such as arrows [11], as well as wider efforts to standardise systems notation [12], medical illustrations, particularly those representing systems diagrams, suffer from a lack of standardisation regarding symbols and nomenclature. This has the potential to result in confusion, with inconsistent interpretations.

As part of a cluster of undergraduate system dynamics courses run for medical and health science students in the Faculty of Health Sciences, we teach system dynamics as a formal computational modelling approach [13,14]. As part of this course, we include causal loop diagrams (CLDs), which are primarily intended to augment the development of the computational models.

CLDs are a simple heuristic that facilitate the representation of a number aspects of a system's behaviour, particularly in the case of systems involving multiple feedback loops and variables. CLDs are often used as an interim step in the production of level-rate diagrams, which are used to produce computer simulations of system behaviour. The objective of this article is to report on an exploratory study to examine the hypothesis that **improved understanding of systems can be achieved by replacing typical medical journal and textbook systems diagrams with their equivalent CLDs.**

The importance of feedback loops and their diagrammatic representation in physiology is well recognised [9]. CLDs are established in the public health literature [15–21], and have been used in other aspects of medical education such as physiology [22]. However, the CLD is a qualitative tool. Its utility lies primarily in making feedback loops explicit and in understanding the relationships between variables. Forrester, the founder of system dynamics, highlights the limitations of CLDs when used without formal quantitative modelling [23]. While acknowledging their value in supporting computational model development and in providing an overall sense of system structure, he cautions against the notion that CLDs alone can capture or explain detailed system behaviour [23].

A number of authors have recognised the importance of systems thinking and causal loop diagrams in improving the understanding of physiological processes [7,24]. However, the definition of CLDs is not always consistently applied. For example, Silldorff et al. describe their diagrams as CLDs and use the effects of antidiuretic hormone as an illustrative example [24], yet their diagrams lack several of the key characteristics that make true CLDs so useful.

What is it about CLDs that make them candidates for potentially replacing typical systems diagrams in journals? One possibility is that CLDs diminish the extraneous load as defined by Cognitive Load Theory [25–27]. Indeed, cognitive load theory has been invoked

in the context of modelling in biology [28]. Their economy of representation reduces clutter and makes the system's interactions more explicit and logically consistent, presenting the student with a direct, unambiguous representation of the system's architecture. The effectiveness of CLDs may also exert an influence through the non-verbal aspect of Paivio's Dual Coding Theory [29].

A Brief Review of CLDs

A CLD comprises only two types of elements: system variables and causal links, shown as arrows with polarity. The system variables are the measurable entities whose behaviour is relevant to the system. The causal arrows are drawn as links between variables to show how one variable affects another.

A high-level CLD drawn in Vensim PLE (Ventana Systems) is shown in Figure 1, representing the treatment of a patient in the healthcare system as an example. A patient may be described in terms of their State of Health (SoH)—a measurable quantity—in principle.

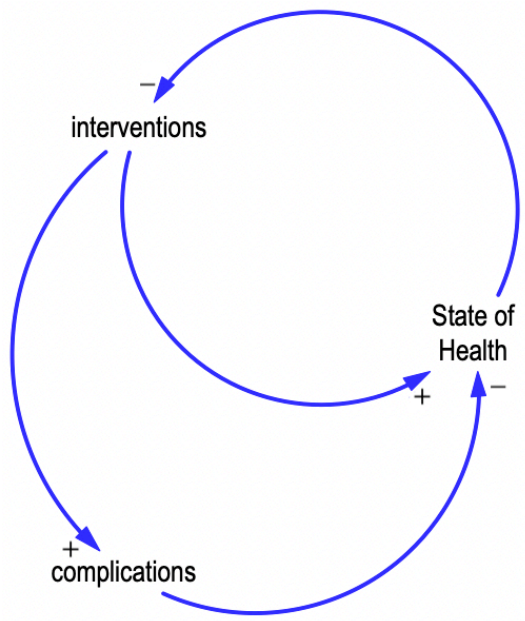

Figure 1. The twin loops of iatrodynamics drawn in Vensim PLE. The negative feedback loop restores the patient's health while the positive (iatrogenic) feedback loop counteracts this.

When the SoH declines, medical professionals generally consider an intervention, which could be any of a number of actions including drugs, surgery, imaging, catheterisation, etc. Ideally, such an intervention, whether diagnostic or therapeutic, will move the SoH back towards normal. This is shown as a causal link from SoH to interventions and from interventions to SoH. The link from SoH to intervention has a "−" polarity symbol, indicating that a change in SoH moves interventions in the opposite direction, i.e., an increase in SoH leads to a decrease in interventions and vice versa. Similarly, the link from interventions to SoH has a "+" polarity symbol, indicating that they move in the same direction, i.e., an increase in interventions leads to an improvement in SoH and vice versa.

However, interventions give rise to complications, and complications reduce the SoH. The occurrence of this adverse outcome in healthcare is described as being iatrogenic, which strictly means complications caused by a healthcare worker or doctor. However, its modern usage is always applied to adverse outcomes. The links for these features of the system are shown in Figure 1. The key insight from this model is that patient care involves two closed loops. The first loop is SoH→interventions→SoH.

The product of the polarities is $(-) \times (+) = (-)$, which indicates that this is a negative feedback (balancing) loop.

Similarly, the loop comprising SoH→interventions→complications→SoH is a positive feedback (reinforcing) loop, sometimes known as a vicious cycle, as the product of the polarities are $(-) \times (+) \times (-) = (+)$.

The outcome of any patient in the healthcare system depends on the balance between these two loops—we could describe the study of these processes as 'iatrodynamics' from the Greek *iatros*, meaning doctor or healer, and *dynamikos*, meaning power or, in the modern context, change. If the negative feedback loop dominates, the patient may recover. However, if the positive feedback loop dominates, it all too frequently results in an inexorable decline and death.

While the term CLD is used as a general descriptor of these drawings, it should be noted that the simple connection of one variable to another through a causal arrow is better described as a causal link. A series of connections is a causal chain, and it is only when the chain feeds back to an earlier variable that it constitutes a causal loop in the strictest sense.

Richardson points out a self-evident but important nuance regarding the effect of the causal link's polarity in CLDs [30]: A positive sign on the arrowhead does not always mean that the responding variable actually moves in the same direction as the causal variable. It may simply mean that if the causal variable increases/decreases, the responding variable will have increased/decreased relative to what it would have been had the causation not acted. A similar argument holds for a negative polarity. For example, the effect of elimination rate of a drug from the circulation on the concentration of drug in the body has a negative causal polarity. However, a declining elimination rate does not necessarily imply a rising concentration of drug, which would also depend on the input rate. Rather, it simply means that the concentration of drug will be higher than it would otherwise have been had the elimination rate not decreased.

A substantial number of articles provide detailed advice on the construction of CLDs such as: [31–36]. While CLDs offer a powerful approach to representing and understanding models, it is important to heed Forrester's caution about their limitations in the absence of formal computational modelling [23]. In particular, CLDs do not inherently show loop dominance. Also, there is no indication as to which loops display short-term and which loops display long-term characteristics. Some authors suggest showing more dominant loops with larger radii [34].

Arguably, one of the most important characteristics of feedback loops is the presence of delays, which have critical consequences for model behaviour [31,36]. In the absence of a quantitative model, it is not possible to infer actual behaviour.

For example, from CLDs alone, it is not possible to infer the existence of periodicity or oscillations in systems such as circadian rhythms, respiratory control, or the menstrual cycle. Oscillations are often attributable to high loop gains or delays. Some authors suggest labelling of causal links that are subject to a delay with the word "Delay" or a corresponding symbol [31,36]. However, while such labelling may alert the reader to the possibility of certain behaviour patterns such as oscillations, the absence of quantification makes it impossible to know this with any certainty.

Correctly drawn CLDs have a number of characteristics [34], which are not typically applied in medical illustrations presented in journals and textbooks. Kim recommends using nouns rather than verbs as variable names [34]. For example, in a model of glucose regulation, it is preferable to have a variable "glucose" rather than "increasing glucose", or as it is so frequently written, "↑glucose", as seen in some literature such as [24]. This avoids confusing use of language, e.g., "increasing insulin causes a decrease in increasing glucose".

It is also essential to avoid repetition of the same variable name in a CLD. Again, this is a common mistake seen in medical texts. While not often articulated, it is also important to maintain generality of CLDs. For example, producing two different CLDs for two pathologies of the same system is inefficient and confusing. A general CLD of the system can be modified to accommodate the various pathologies, providing far more insight.

2. Materials and Methods

Two journal-based endocrine systems diagrams were selected for this study from a range of online options that met predefined criteria, namely publication in well-recognised, peer-reviewed, open-access journals, coverage of the same broad topic (in this case, endocrinology) that are typical of journal illustrations depicting systems, and comparable levels of complexity—the selected diagrams comprise six and nine feedback loops respectively.

The suitability of the diagrams based on the above criteria were determined by three of the authors who are domain experts in that they teach the system dynamics course and hold medical and engineering qualifications (two holding both). The selected images exhibit several features that causal loop diagrams are designed to address, and we therefore judged them to be appropriate exemplars. Future work may benefit from the development of a more formal coding framework for illustration selection, incorporating measures of image complexity and structural error.

The equivalent CLDs were constructed using Vensim® PLE, Version 10.4.0 (Ventana Systems, Inc., Harvard, MA, USA) software. The same three authors reviewed the CLDs to ensure that their architectures faithfully represent the system structure depicted in the journal diagrams.

The first is a pair of diagrams presented as Figure 3 (https://academic.oup.com/view-large/figure/367629299/eje-19-0316fig3.jpeg accessed on 29 March 2026) and Figure 4 (https://academic.oup.com/view-large/figure/367629321/eje-19-0316fig4.jpeg accessed on 29 March 2026) in reference [37], representing two specific clinical scenarios of calcium homeostasis, namely chronic hypoparathyroidism and primary hyperparathyroidism. Each of the two journal diagrams of the calcium homeostasis system represent different pathologies. Note that the journal diagrams can be accessed by clicking on the hyperlinks above.

The first figure of the pair of journal diagrams shows the system in the presence of parathyroid gland failure as, for example, may occur as a complication of thyroid surgery. The second figure in the pair of diagrams represents the system in the context of a benign parathyroid tumour that secretes excess parathyroid hormone (PTH) irrespective of the serum calcium level.

Both cases represent an open-loop condition as the feedback from serum ionised calcium to PTH is broken. In the first pathology, the serum calcium is low. But in the second case, it is excessively high.

We treated both journal diagrams of calcium pathology as a single entity and combined them into a single CLD shown in Figure 2, drawn in Vensim PLE (Ventana Systems). We re-introduced the causal link between ionised calcium and PTH in the CLD, thus reintroducing closed loop control. This is in keeping with the underlying philosophy of systems representations where the CLD is generic and can be adapted for any specific scenario—it is up to the student to remove the link in the case of an open loop system.

The second journal diagram is of glucose homeostasis presented as Figure 1 (https://www.frontiersin.org/files/Articles/928016/fendo-13-928016-HTML/image_m/fendo-13-928016-g001.jpg accessed on 29 March 2026) in reference [38]. It includes the paracrine interaction of insulin and glucagon, as well as the effects on glycogen uptake, storage, glycogenolysis, and gluconeogenesis. Again, the journal diagram can be viewed by clicking on the hyperlink above.

A CLD, drawn in Vensim PLE (Ventana Systems), was created for the journal diagram depicting glucose homeostasis as shown in Figure 3.

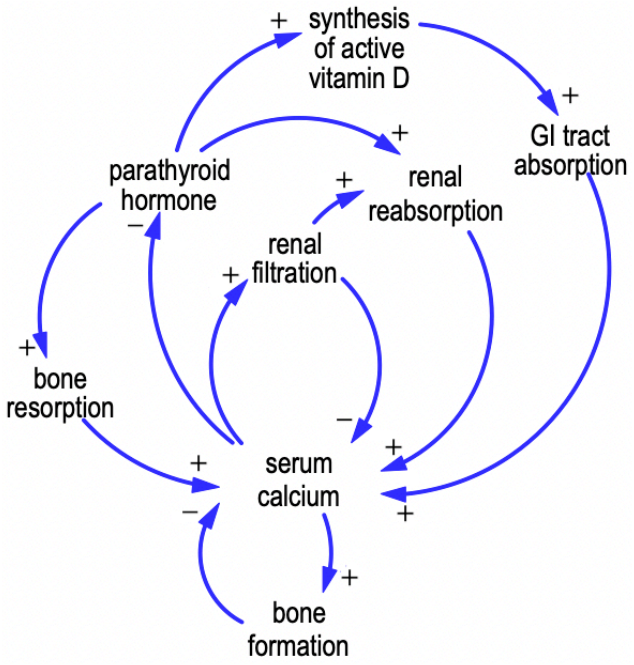

Figure 2. CLD version of calcium homeostasis model from Figures 3 and 4 in reference [37] to regulate blood calcium concentration. The diagram comprises six feedback loops—five balancing loops and one reinforcing loop.

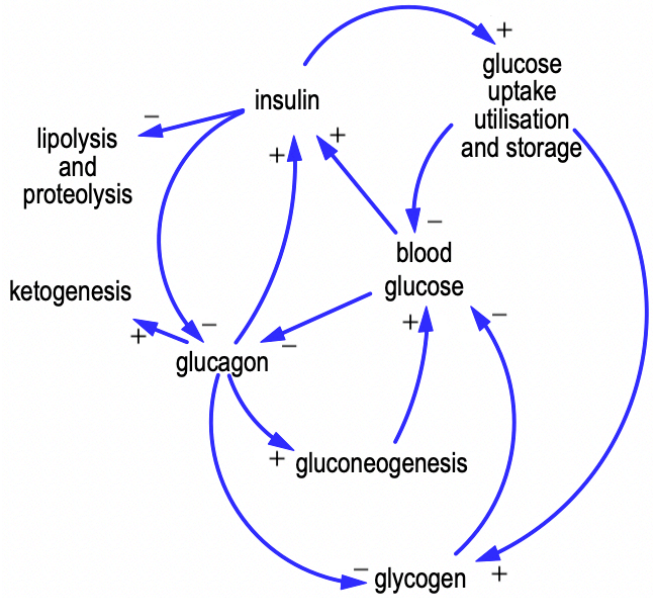

Figure 3. CLD version of the glucose homeostasis model presented as Figure 1 in reference [38] to regulate blood glucose concentration. The diagram comprises seven negative and two positive feedback loops.

Two tests were created for the study.

Test 1 comprised both journal diagrams of calcium homeostasis on one page, which included both their figure captions, directly from the journal. Both diagrams were presented on the page and treated as a single entity. The CLD of glucose homeostasis was presented on the next page.

Test 2 comprised the CLD for calcium homeostasis on the first page of the test, followed by the journal diagram of glucose homeostasis including its figure caption on the next page.

In both tests, this was followed by 12 true/false questions, 6 for each of the two systems. While true/false questions have limitations in assessing deeper understanding, their use in this study served to minimise ambiguity in responses, allow rapid assessment under strict time constraints, and provide objective, easily quantifiable data suitable for statistical analysis.

The questions were identical for both tests and are shown below. In this way, each test had a CLD and a journal diagram as shown in Table 1. This is a quasi-experimental, counterbalanced, mixed design. The mixed design allowed assessment of within-subject effects of system and between-subject effects of diagram assignment.

Table 1. Composition of the two tests.

	Test 1	**Test 2**
Calcium	Journal diagram	CLD
Glucose	CLD	Journal diagram

Calcium homeostasis questions

Indicate TRUE or FALSE at the end of each statement (A) to (F). Please carefully look at the associated drawing(s) to answer each question.

(A) A complete negative feedback (goal-seeking) loop is formed by just the two variables, **renal filtration** and **serum calcium (Ca^{2+})**.

(B) The three variables, **bone resorption, serum calcium (Ca^{2+})**, and **parathyroid hormone (PTH)**, form a complete negative feedback (goal-seeking) loop.

(C) A patient can no longer produce any **parathyroid hormone (PTH)** due to surgical damage to the parathyroid glands. TRUE or FALSE: In this situation, **GI-tract absorption** is no longer part of any feedback loop.

(D) A tumour in a patient's parathyroid gland produces large quantities of **parathyroid hormone (PTH)**, which is not sensitive to **serum calcium (Ca^{2+})**. In other words, the **serum calcium (Ca^{2+})** no longer influences the production of **parathyroid hormone (PTH)**. TRUE or FALSE: In this situation, **GI-tract absorption** is no longer part of any feedback loop.

(E) A complete negative feedback (goal-seeking) loop is formed by just the two variables, **bone formation** and **serum calcium (Ca^{2+})**.

(F) The variables **parathyroid hormone (PTH)**, **renal reabsorption**, and **GI-tract absorption** constitute a valid feedback loop.

Glucose homeostasis question

Indicate TRUE or FALSE at the end of each statement (A) to (F). Please carefully look at the associated drawing to answer each question.

(A) The variables, **blood glucose**, **glucagon**, and **glycogen**, constitute a positive feedback (reinforcing) loop.

(B) The variables, **glycogen**, **blood glucose**, **insulin**, and **glucagon**, constitute a valid negative feedback (goal-seeking) loop.

(C) The variables **blood glucose**, **glucagon**, and **gluconeogenesis** constitute a valid positive feedback (reinforcing) loop.

(D) The variables **insulin**, **glucose uptake utilisation and storage**, and **blood glucose** constitute a valid negative feedback (goal-seeking) loop.

(E) Decreased **blood glucose** leads to a decrease in **glycogen**.

(F) Increased **insulin** results in a decrease in **lipolysis and proteolysis**.

It was not possible to test a cohort who had not been exposed to causal loop diagrams, as there is no reasonable prospect that students who are not familiar with the nomenclature of CLDs could interpret them correctly.

Test administration: Participants from the first year course in Health System Dynamics in the Bachelor of Health Science programme were recruited in terms of the ethics approval (Protocol number: M250622 as amended on 22 August 2025) from our institutional review board, HREC (Medical). The study was conducted during lunch hour and participants were compensated with a modest lunch voucher for an on-campus vendor, as required in terms of our ethics approval. The cohort is entirely comprised of first-year health science students who are mostly premed. The course is essentially the same as the original prototype course developed for the 6-year medical degree [13].

Based on prior expression of interest and typical drop-out rates, an unexpectedly large group of 55 students arrived to participate in the study. As we considered group size parity to be important for a small group, and as participant numbers were unknown until they arrived for the study, it was not possible to conduct prior randomisation using a method such as block allocation to ensure strict randomisation while maintaining group size parity.

Thus, participants were allocated to Test 1 (28 students) or Test 2 (27 students) using alternating assignment based on their self-selected seating order. This approach was chosen for its practicality in a classroom setting under severe time constraints to ensure balanced group sizes, which, as mentioned above, was a priority given the initially anticipated smaller sample. Furthermore, any local clustering of participants (e.g., by ability) would be distributed across tests by the alternating pattern, reducing the risk of systematic bias. As seating was not assigned, and the individuals administering the tests had no personal knowledge of student ability in this subset, which was drawn from a very large class, the risk of systematic bias was minimal. Given the counterbalanced design and the absence of any *a priori* expectation of bias, this approach was considered adequate for this exploratory study.

Analysis: A two-way repeated measures ANOVA using a Mixed Linear Model was conducted to examine the effects of representation format type, i.e., CLD or journal diagram, and the specific system being depicted, either Calcium or Glucose, on participant performance. The model treated the system as a within-subjects factor and the diagram presentation method as a between-subjects factor. The fit was achieved using the SAS PROC MIXED (SAS/STAT®, Version 15.2, NC, USA) procedure. This will be further elaborated upon in the Section 3.

A Fisher's Exact Test was performed to compare the individual questions for **CLD** vs. **JSD** for each system.

3. Results

For the purpose of examining the statistical features of the results, it is necessary to clearly define six terms. The drawings fall into two categories, namely **CLD** and Journal System Drawing (**JSD**). Collectively, both categories will be known as **Diagram(s)**. The drawings are of two different endocrine systems and we will call these **Calcium** and **Glucose**, respectively. Collectively, these two endocrine systems will be known as **System(s)**. Whenever these terms are used in this context, they will be capitalised and bolded. Thus, there are four categories of drawings, shown in the format of **System, Diagram**:

Calcium, CLD
Calcium, JSD
Glucose, CLD
Glucose, JSD

The effects on participant performance, of representation format (**Diagram**)—either **CLD** or **JSD**—and the specific system being depicted (**System**)—either **Calcium** or **Glucose**—was examined by two-way repeated measures ANOVA using a Mixed Linear Model. The model treated System as a within-subjects factor and drawing presentation method as a between-subjects factor, and it was fit using the SAS PROC MIXED procedure. Although linear models assume normally distributed error terms and the individual items comprise sums of binary responses, the use of a linear model is justified by the Central Limit Theorem, which applies to the aggregation of responses both within the scale itself—even though the items are not strictly independent—and across observations. A histogram of the residuals confirms an approximately normal distribution.

The ANOVA summary (Table 2) shows significant main effects of **System**, $F(1, 53.48) = 12.39$, $p < 0.001$, and **Diagram**, $F(1, 53.77) = 125.76$, $p < 0.001$, as well as a significant interaction between **System** and **JSD**, $F(1, 53.48) = 20.11$, $p < 0.001$. These results indicate that both the representation format (**Diagram**) and the **System** independently affected performance, and the effect of **Diagram** differed depending on the **System** being tested.

Table 2. ANOVA summary for the effects of **System**, **Diagram**, and their interaction is shown in terms of degrees of freedom (df), the F statistic and the *p*-value.

Effect	df	F	*p*
Diagram	1, 53.48	125.76	<0.001
System	1, 53.77	12.39	<0.001
Diagram × **System**	1, 53.48	20.11	<0.001

As the interaction was significant, post hoc comparisons with a Bonferroni adjustment were conducted to examine the simple effects of representation format (Diagram) within each System. Least Squares Means revealed that for the Calcium questions, scores were substantially higher (mean difference = 0.46, SE = 0.041, 95% CI [0.38, 0.54]) when presented as a CLD (mean = 0.84), than when presented as the JSD (mean = 0.38). For the Glucose questions, the same pattern was observed, although the difference was smaller (mean difference = 0.20, SE = 0.041, 95% CI [0.11, 0.28]): scores were higher for CLD (mean = 0.83) compared to the JSD (mean = 0.63). These results, expressed as percentages, are summarised in Table 3, which highlight the magnitude of the effect sizes.

Table 3. Interaction table for **Diagram** × **System**, showing Least Squares Means, Standard Error, and 95% confidence intervals (CIs), all expressed as percentages.

				95% CI	
System	**Diagram**	**Least Squares Means**	**Standard Error**	**Lower**	**Upper**
Calcium	**CLD**	84.01%	3.18%	77.78%	90.24%
Calcium	**JSD**	38.10%	3.12%	31.97%	44.23%
Glucose	**CLD**	82.74%	3.12%	76.61%	88.87%
Glucose	**JSD**	63.05%	3.18%	56.82%	69.28%

The interaction plot (Figure 4) shows that while **CLD** yields similarly high scores for both **Calcium** and **Glucose**, performance under the **JSD** representation drops sharply for **Calcium** and moderately for **Glucose**. Thus, the advantage of the CLD method is much greater for **Calcium** than for **Glucose**, confirming a strong **Diagram** × **System** interaction.

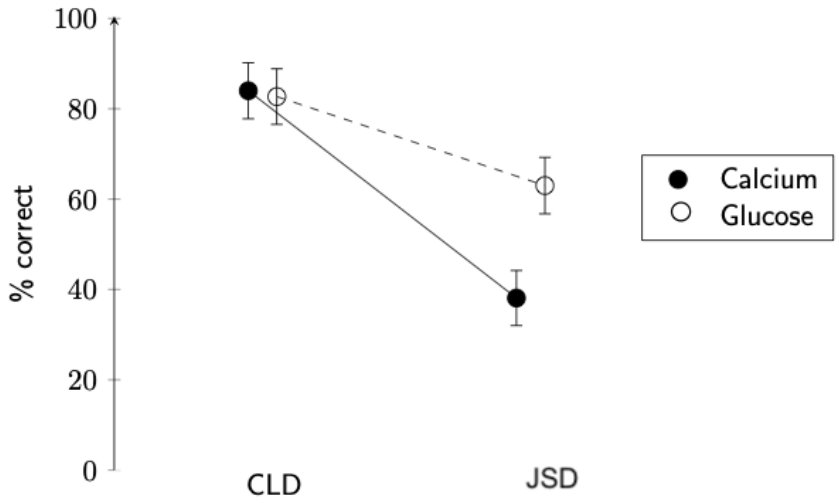

Figure 4. Interaction plot of **Diagram** × **System**, where **Diagram** refers the representation format (**CLD**, **JSD**) and **System** is the model type (**Calcium**, **Glucose**). The error bars represent the 95% confidence intervals.

The Fisher's Exact Test for the **Calcium** system showed higher proportions of correct answers for all the questions for the **CLD** compared to **JSD**, but in one of them, question D, participants performed poorly on both **Diagram** types with no statistically significant difference. These results are shown in Table 4 and Figure 5.

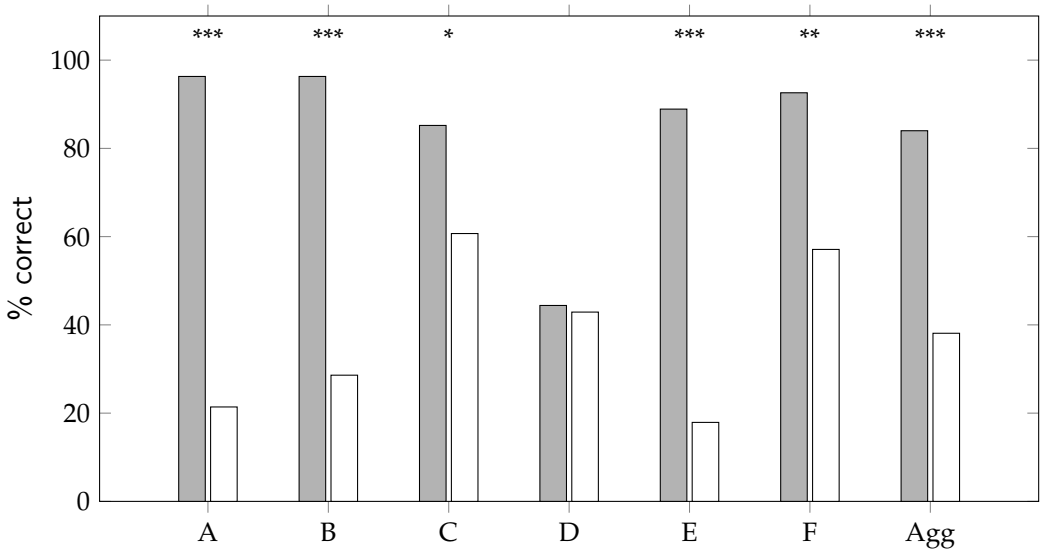

Figure 5. Bar graphs for calcium homeostasis showing % correct for **CLD** (grey) and **JSD** (white) (Figures 3 and 4 from [37]) for questions $A \rightarrow F$, as well as the aggregate of all questions $\bigcup_A^F$ (Agg). Statistical significance on Fisher's Exact Test: $p \leq 0.05$ (*), $p \leq 0.01$ (**), and $p \leq 0.001$ (***).

Table 4. Results of Fisher's Exact Test for the calcium homeostasis questions $A \rightarrow F$, as well as the aggregate of all questions $\bigcup_A^F$.

	A	**B**	**C**	**D**	**E**	**F**	$\bigcup_A^F$
Correct (%)							
CLD	96.3	96.3	85.2	44.4	88.9	92.6	84
JSD [37]	21.4	28.6	60.7	42.9	17.9	57.1	38.1
p one-tail (right)	$\ll 10^{-4}$	$\ll 10^{-4}$	0.04	0.6	$\ll 10^{-4}$	0.003	$\ll 10^{-4}$
Odds Ratio	95	65	3.7	1.1	36.8	9.4	8.5
95% CI	11–853	8–563	1–14	0.4–3	8–172	2–48	5–14

On the glucose model, all questions other than B showed a higher proportion of correct answers on **CLD** than on **JSD**. However, only in questions C and F were the differences statistically significant. These results are shown in Table 5 and Figure 6.

For the **Glucose** model, while the proportion of correct answers was greater for **CLD** compared to **JSD** for all but one of the questions, only two questions showed a statisitically significant difference.

Table 5. Results of Fisher's Exact Test for the glucose homeostasis questions $A \rightarrow F$, as well as the aggregate of all questions $\bigcup_A^F$.

	A	**B**	**C**	**D**	**E**	**F**	$\bigcup_A^F$
Correct (%)							
CLD	89.3	67.9	92.9	92.9	60.7	92.9	83
JSD [38]	70.4	74.1	44.4	81.5	55.6	51.9	63
p one-tail (right)	0.08	0.8	10^{-4}	0.2	0.45	0.0007	$<10^{-4}$
Odds Ratio	3.50	0.74	16.25	2.95	1.24	12.07	2.82
95% CI	0.8-15	0.2–2.4	3.2–83	0.5–17	0.4–3.6	2.4–61	1.7–4.7

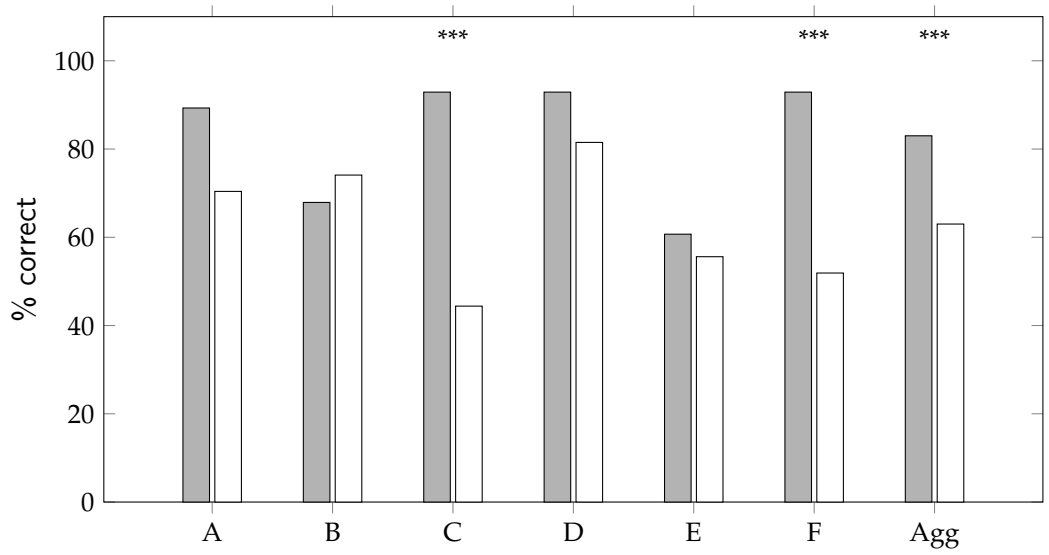

Figure 6. Bar graphs for glucose homeostasis showing % correct for **CLD** (grey) and **JSD** (white) (Figure 1 from [38]) for questions $A \rightarrow F$, as well as the aggregate of all questions $\bigcup_A^F$ (Agg). Statistical significance on Fisher's Exact Test: $p \leq 0.05$ (*), $p \leq 0.01$ (**), and $p \leq 0.001$ (***).

4. Discussion

Clear and efficient diagrammatic representations are essential to systems thinking. The medical literature has an abundance of diagrams representing systems of varying degrees of complexity, yet there is currently no standard symbolic nomenclature, giving rise to the potential for confusion and misunderstanding.

The well-established conventions for CLDs have the potential to reform medical systems diagrams. We formulated the hypothesis that improved understanding of systems can be achieved by replacing typical medical journal and textbook diagrams with their equivalent CLDs. We conducted an exploratory examination of this hypothesis by selecting two sets of drawings from the peer-reviewed medical literature. We reconstituted these drawings as CLDs, taking care to ensure that the CLDs exactly represented the journal drawings. We presented both drawing types for both systems, unseen, in a counterbalanced study design, to health science student volunteers.

Both of the journal drawings had inconsistencies and features that are specifically avoided in CLDs. Both systems had variables that were represented as verbs rather than

nouns in the journal diagrams, for example, ↓**reabsorption** and **blood glucose**↑. The journal diagram of glucose homeostasis had glucose repeated twice, once for increase and once for decrease, and it included as variables items that did not represent measurable quantities in the context of this system, namely α and β cells. It should be emphasised that while it is fully acceptable to show anatomical structures on the diagram to indicate where these processes are taking place, they should not be represented as variables in the system diagram unless they are explicitly modelled as such—for example, in the case where the β cell mass is being modelled.

The calcium system showed two pathologies that were represented by two separate systems diagrams in the journal drawings. In both cases, there was no feedback from ionised calcium to the PTH production. These could have easily been combined into a single drawing, as we did with the equivalent CLD where we left the feedback in place, and expected the participants to be able to recognise the failure of feedback and modify the drawing accordingly. Doing so introduces generalisation into the drawing, making it adaptable to the various pathologies that affect this system.

The two-way repeated measures ANOVA indicates that participant performance was influenced by both the type of diagram (CLD vs. JSD) and the system being depicted (Calcium vs. Glucose). Participants showed substantially better performance on the CLDs than on the JSDs across both systems (Calcium and Glucose). However, their performance on the journal drawings were not equal. Participants achieved substantially higher scores on the journal diagram for glucose homeostasis than on the calcium system.

The CLD approach yields improvements of 46% for the calcium diagram and 20% for the glucose diagram. These are substantial improvements, particularly relative to the standard deviations of their respective groups (27% and 20%). Although the sample sizes are relatively small in both cases, the lower bounds of the confidence intervals (38% and 11%, respectively) still indicate meaningful improvements in the corresponding test scores.

The reason for this difference between the calcium and glucose systems is uncertain. It could be due to unintended complexity introduced into the calcium system compared to the glucose system by representing it as two separate diagrams—a strong motivation to resort to CLDs. It is also possible that as glucose homeostasis is a well-known system due to the prevalence of diabetes, students may already be familiar with the system. Further investigation of these possibilities is warranted.

Interestingly, for the calcium system, while students performed far better on the CLD than on the journal diagram in question C, their performance on question D was statistically identical between diagram types. Question C relates to failure of the parathyroids, whereas question D relates to an autonomously secreting parathyroid tumour.

Both pathologies represent open-loop systems due to failure of the feedback. In the first case (question C), the damaged parathyroid glands do not respond to calcium and thus do not produce PTH. In the second case (question D), one of the parathyroid glands produces excessive PTH but it is not responsive to calcium levels, having lost sensitivity due to aberrant characteristics of the tumour. Participants recognised the first case as an open-loop control but not the second case. This suggests a possible teaching intervention using other examples such as autonomous thyroid or secretory pituitary tumours.

For three of the glucose questions, A, B, and E, there was no statistically significant difference between the CLD and the JSD.

Study Limitations

This exploratory study has a number of limitations. The study was limited to two examples from journal articles. While this was partially mitigated by the counterbalanced design, where each system served as its own control, the unequal performance of students

in the two system types suggests that an expansion of this study to include a wider range of examples may be illuminating. Broadening the scope of this study to include a greater diversity of examples would likely generate deeper analytical insights and improve the generalisability of the results. This extension would facilitate a more systematic and reproducible analytical framework. This could be achieved through the development of a structured coding scheme that classifies illustrations based on specific attributes.

These attributes could encompass the type of system depicted, the presence of particular symbols, and the overall structural complexity of the system. Complexity, in particular, could be quantified using metrics such as the number or density of feedback loops, as well as the hierarchical organisational level. Implementing a standardised coding protocol across a sufficiently large sample would facilitate more rigorous comparative analyses while enhancing reproducibility.

Future studies would potentially benefit from studying larger cohorts at different stages of their curriculum, as well as from more stringent randomisation. This would be expected to increase generalisability and improve confidence in the results.

Finally, the single-institution, single-cohort design and the use of true/false questions could limit the information yield of this study and may be addressed in future work.

5. Conclusions

Medical illustrations vary widely in quality, and there is no standardisation of nomenclature or symbols. This exploratory study examines the potential of CLDs to enhance medical students' understanding of the complex systems they encounter in their curriculum by replacing traditional journal and textbook drawings.

It seems plausible that the effectiveness of CLDs over typical journal drawings may be associated with the lowering of extraneous load [27], and further study on the mechanistic aspects is warranted. These preliminary findings suggest that exposure to CLDs may have the potential to enhance systems thinking and understanding among medical and pre-medical students. However, further investigation is needed to more comprehensively address a number of limitations including the paucity of diagram types and the relatively small, single-institution sample.

Author Contributions: D.M.R. proposed, initiated, led the development, and taught the first course in system dynamics for the medical school. X.L.R. was involved in the development and teaching of the course from its very inception. D.M.R. conceived, designed, and led this study together with S.A. and X.L.R. S.A. currently coordinates and teaches the cohort from which the study participants were drawn. X.L.R., S.A. and A.P. further developed and taught versions of the course, which all contributed to the intellectual corpus of material on which this study draws. A.G. provided the pedagogical, ethical, and organisational understanding that shaped the study. D.M.R. wrote the first draft of the manuscript, which was further developed by all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding; all financial support was provided through internal institutional research funds.

Informed Consent Statement: Participants gave their informed consent in terms of ethics approval, Protocol M250622, as amended on 22 August 2025 from HREC (Medical).

Data Availability Statement: The anonymised raw data will be made available upon reasonable request.

Acknowledgments: We thank data scientist, Michael Greyling, for suggesting and running the ANOVA and for confirming our results on the Fisher Exact Test.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bouras-Vallianatos, P. Diagrams in Greek Medical Manuscripts. In *The Diagram as Paradigm: Cross-Cultural Approaches*; Hamburger, J.F., Roxburgh, D.J., Safran, L., Eds.; Dumbarton Oaks Research Library and Collection: Washington, DC, USA, 2022; pp. 287–329.
2. Müller, H.; Kalpathy-Cramer, J.; Demner-Fushman, D.; Antani, S. Creating a classification of image types in the medical literature for visual categorization. In *SPIE Medical Imaging, Proceedings of the Medical Imaging 2012: Advanced PACS-Based Imaging Informatics and Therapeutic Applications, San Diego, CA, USA, 4–9 February 2012*; SPIE: San Diego, CA, USA, 2012; Volume 8319, pp. 195–206.
3. Guerra, D. What type of medical illustrations best improve understanding of scientific content? A systematic review. *J. Vis. Commun. Med.* **2025**, *48*, 21–43. [CrossRef]
4. Tippett, C.D. What recent research on diagrams suggests about learning with rather than learning from visual representations in science. *Int. J. Sci. Educ.* **2016**, *38*, 725–746. [CrossRef]
5. Wilson, K.J.; Long, T.M.; Momsen, J.L.; Bray Speth, E. Modeling in the classroom: Making relationships and systems visible. *CBE—Life Sci. Educ.* **2020**, *19*, fe1. [CrossRef]
6. Momsen, J.; Speth, E.B.; Wyse, S.; Long, T. Using systems and systems thinking to unify biology education. *CBE—Life Sci. Educ.* **2022**, *21*, es3. [CrossRef]
7. Murphy, M.P.; Hung, W. Using systems modeling to facilitate undergraduate physiology student learning and retention of difficult concepts. *Adv. Physiol. Educ.* **2024**, *48*, 867–872. [CrossRef] [PubMed]
8. Scott, E.E.; Anderson, C.W.; Mashood, K.; Matz, R.L.; Underwood, S.M.; Sawtelle, V. Developing an analytical framework to characterize student reasoning about complex processes. *CBE—Life Sci. Educ.* **2018**, *17*, ar49. [CrossRef] [PubMed]
9. Wellmanns, A.; Schmiemann, P. Feedback loop reasoning in physiological contexts. *J. Biol. Educ.* **2022**, *56*, 465–485. [CrossRef]
10. Kottmeyer, A.M.; Van Meter, P.; Cameron, C. Diagram comprehension ability of college students in an introductory biology course. *Adv. Physiol. Educ.* **2020**, *44*, 169–180. [CrossRef] [PubMed]
11. Wright, L.K.; Cardenas, J.J.; Liang, P.; Newman, D.L. Arrows in biology: Lack of clarity and consistency points to confusion for learners. *CBE—Life Sci. Educ.* **2017**, *17*, ar6. [CrossRef]
12. Novère, N.L.; Hucka, M.; Mi, H.; Moodie, S.; Schreiber, F.; Sorokin, A.; Demir, E.; Wegner, K.; Aladjem, M.I.; Wimalaratne, S.M.; et al. The systems biology graphical notation. *Nat. Biotechnol.* **2009**, *27*, 735–741. [CrossRef]
13. Rubin, D.M.; Richards, C.L.; Keene, P.A.; Paiker, J.E.; Gray, A.R.; Herron, R.F.; Russell, M.J.; Wigdorowitz, B. System dynamics in medical education: A tool for life. *Adv. Health Sci. Educ.* **2012**, *17*, 203–210. [CrossRef]
14. Rubin, D.M. Inclusion of system dynamics training in the education of healthcare workers in Africa. In *System Dynamics Models for Africa's Developmental Planning*; Brent, A., Simelane, T., Eds.; Africa Institute of South Africa: Pretoria, South Africa, 2018; pp. 213–234.
15. Sterman, J.D. Learning from evidence in a complex world. *Am. J. Public Health* **2006**, *96*, 505–514. [CrossRef]
16. Tomoaia-Cotisel, A.; Kim, H.; Allen, S.D.; Blanchet, K. 16 Causal loop diagrams: Emergent system behaviour. In *Applied Systems Thinking for Health Systems Research: A Methodological Handbook*; Open University Press: Berkshire, UK, 2017; p. 97.
17. Baugh Littlejohns, L.; Baum, F.; Lawless, A.; Freeman, T. The value of a causal loop diagram in exploring the complex interplay of factors that influence health promotion in a multisectoral health system in Australia. *Health Res. Policy Syst.* **2018**, *16*, 126. [CrossRef] [PubMed]
18. De Pinho, H. Generation of systems maps. In *Systems Science and Population Health*; Oxford University Press: New York, NY, USA, 2017; p. 61.
19. Lembani, M.; de Pinho, H.; Delobelle, P.; Zarowsky, C.; Mathole, T.; Ager, A. *A Guide for Participatory Systems Analysis Using a Group Model Building Approach*; SAGE Publications Limited: London, UK, 2020.
20. Waterlander, W.E.; Singh, A.; Altenburg, T.; Dijkstra, C.; Luna Pinzon, A.; Anselma, M.; Busch, V.; van Houtum, L.; Emke, H.; Overman, M.L.; et al. Understanding obesity-related behaviors in youth from a systems dynamics perspective: The use of causal loop diagrams. *Obes. Rev.* **2021**, *22*, e13185. [CrossRef]
21. Uleman, J.F.; Stronks, K.; Rutter, H.; Arah, O.A.; Rod, N.H. Mapping complex public health problems with causal loop diagrams. *Int. J. Epidemiol.* **2024**, *53*, dyae091. [CrossRef]
22. Rubin, D.M.; Letts, R.F.R.; Richards, X.L. Teaching physiology within a system dynamics framework. *Adv. Physiol. Educ.* **2019**, *43*, 435–440. [CrossRef]
23. Forrester, J.W. System dynamics, systems thinking, and soft OR. *Syst. Dyn. Rev.* **1994**, *10*, 245–256. [CrossRef]
24. Silldorff, E.P.; Robinson, G.D. How and why to focus on causality in teaching human physiology. *Adv. Physiol. Educ.* **2025**, *49*, 1094–1101. [CrossRef] [PubMed]
25. Sweller, J.; Chandler, P. Evidence for cognitive load theory. *Cogn. Instr.* **1991**, *8*, 351–362. [CrossRef]
26. Paas, F.; Renkl, A.; Sweller, J. Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* **2003**, *38*, 1–4. [CrossRef]
27. Sweller, J. Cognitive load theory. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 55, pp. 37–76.

28. Dunbar, T. Supporting Feedback Loop Reasoning in Simulated Systems with Computer-Based Scaffolding. Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2023.
29. Clark, J.M.; Paivio, A. Dual coding theory and education. *Educ. Psychol. Rev.* **1991**, *3*, 149–210. [CrossRef]
30. Richardson, G.P. Problems in causal loop diagrams revisited. *Syst. Dyn. Rev. J. Syst. Dyn. Soc.* **1997**, *13*, 247–252. [CrossRef]
31. Kim, D.H.; Anderson, V. *Systems Archetype Basics*; Pegasus Communications Inc.: Waltham, MA, USA, 1998.
32. Kim, D. *Introduction to Systems Thinking*; Pegasus Communications: Waltham, MA, USA, 1999; Volune 16.
33. Ford, A. *Modeling the Environment*, 2nd ed.; Island Press: Washington, DC, USA, 2010.
34. Kim, D. Guidelines for drawing causal loop diagrams. *Systems Thinker* **2011**, *22*, 5–7.
35. Sterman, J. Fine-Tuning Your Causal Loop Diagrams—Part I. *Systems Thinker* **2001**, *12*, 6–7.
36. Sterman, J. Fine-Tuning Your Causal Loop Diagrams—Part II. *Systems Thinker* **2001**, *12*, 6–7.
37. Bollerslev, J.; Schalin-Jäntti, C.; Rejnmark, L.; Siggelkow, H.; Morreau, H.; Thakker, R.; Sitges-Serra, A.; Cetani, F.; Marcocci, C.; the PARAT Workshop Group. Unmet therapeutic, educational and scientific needs in parathyroid disorders. *Eur. J. Endocrinol.* **2019**, *181*, 1–19. [CrossRef] [PubMed]
38. Jia, Y.; Liu, Y.; Feng, L.; Sun, S.; Sun, G. Role of Glucagon and Its Receptor in the Pathogenesis of Diabetes. *Front. Endocrinol.* **2022**, *13*, 928016. [CrossRef] [PubMed]